Politecnico di Torino

Master of Science Course in
Cinema and Media Engineering
A.a. 2023/2024
Sessione di Laurea Dicembre 2024

Psychoacoustic Assessment on off-road Vehicles

Supervisors:
Arianna Astolfi
Louena Shtrepi

Candidate:
Elena Forneris

Index

1.Introduction

In the agricultural world, tractors are an indispensable tool for farmers' work. The characteristics and requirements of these machines are increasingly stringent from a technological and performance point of view, with means capable of facing the most difficult territories and for very long working hours.

From a technological point of view, the tractors have made significant progress. The cabin of agricultural vehicles has also undergone many technological innovations, both regarding instrumentation and in terms of comfort and well-being for the user. These improvements, however, are still at a rather late stage as regards the acoustic comfort inside the tractor cab. From the acoustic point of view, investigations and studies have been conducted mainly on external noise, the engine and the vibrations, squeaks and internal noises it causes. The cab is therefore being insulated and there is already an interest in the acoustic comfort of tractor users.

What is not considered, however, is the fact that farmers work in very hot conditions and for extended hours, which forces them to seek relief by using HVAC, commonly called air conditioning. HVAC is a fundamental component of every vehicle, as it allows to have a high comfort from the point of view of the temperature. This is particularly true for agricultural vehicles such as tractors, since the working conditions to which users are subjected are very heavy from a temperature point of view.

The HVAC block creates a noise level inside the cab, however, which is harmful to farmers who spend hours on the trucks, as exposure to too high levels of noise for too long are a risk to the physical and psychological health of users, with short- and long-term damage not to be underestimated.
This thesis, therefore, investigates the acoustic quality inside a tractor, with objective measurements and tests subjective, the results of which have been combined and used to calculate a linear predictive noise model same. The research is the continuation of the thesis work carried out by my colleague Marco Favaretto [39]. Many articles and methodologies are the same as those used by him. On the other hand, they have been differences in methodology and insights at various stages of work. Favaretto's work will often be cited, both as a source and for comparison with acquired data, in order to assess whether the differences introduced from a methodological point of view have been effective. The methodology consists of several parts. The methodology is multi-part.

Initially, a literature review was conducted on the topic to have a verified methodology, to implement a choice of sound and to choose an effective prediction model and a correct strategy for conducting subjective tests. This first part allowed to have a clear picture of the state of art and research in the field.

The second phase consists of the acquisition of signals in the tractor cab, with the various settings studied by the HVAC block. Three different types of microphones were used for this part, such as a 19-channel microphone, an omnidirectional sound meter and a binaural microphone consisting of an artificial head. The measurements were carried out with the engine off, with the tractor located outdoors in a quiet area, with the cab doors closed, to simulate the actual use and the situation of use. The microphones were therefore placed at the driver's seat about the height where the driver's ears would be. The recordings were essential to extract the parameters and objective values of the noise produced by the HVAC block. The psycho-acoustic parameters analyzed and extracted were varied, all supported by literature, such as loudness, sharpness, roughness, tonality, fluctuation strength, A-weighted sound pressure level and non-weighted sound pressure level. This allowed a very clear and wide picture of the behavior and type of noise in the cab.

The third stage consists of subjective tests, which are essential to have real human perception of noise. The tests were conducted with three different methodologies: a first test was carried out on the day of measurements, with the subjects tested in the tractor cab listening to the noise produced by the HVAC in person. In a separate location, three replaying tests were carried out, one binaural in headphones, the second one in the Audio Space Lab in Politecnico of Turin, with a third order ambisonics system of audio reproduction ad a VR headset, and the other with the aid of a VR headset and a set of headphones reproducing spatialized audio, where the recordings detected on the ground were used as tracks. In order to limit the statistic variance of the results, the laboratory tests were repeated by 5 subjects. The tests asked subjects two types of questions: a scale from 1 to 10 which assessed the annoyance of sound, the other following the method of semantic differential methods, studied in literature, with semantic scales relating to all the objective psycho-acoustic parameters studied.
The predictive capabilities of the model are rigorously assessed, with particular attention to its accuracy in forecasting psychoacoustic parameters and subjective noise ratings. Preliminary findings suggest promising

outcomes, with the 1-10 rating scale demonstrating exceptional efficacy in predicting noise annoyance, achieving an impressive R-squared value of 0.99. Additionally, the Semantic Differential Method (SDM) showcases its utility in predicting psychoacoustic parameters, with no R-squared value lower than 0.88. Moreover, annoyance predictions calculated with spectrum-based parameters reported even more impressive results.
However, challenges are encountered in reliably predicting sharpness, attributed to significant errors and discrepancies in subjective ratings. The findings of this study form the basis for the development of advanced predictive models, potentially using neural networks. Compared to the work of my colleague Favaretto [39], was introduced to deepen the validity of linear regression in the cab of the tractor, which brought valid results and interest for research, confirming it a useful choice for research purposes. In addition, a further step was introduced, free from bibliographical research, such as a methodology for translation of psycho-acoustic parameters from English to Italian with the help of three native and bilingual subjects. This was done to have a semantic scale and terms as clear and precise as possible, so that they can be submitted to an audience of native Italian speakers without losing the semantic nuances of English terminology.

The aim of this research is to develop a study methodology which will improve the acoustic conditions in off-road vehicle cabins, thus improving the quality of work for farmers. Finally, the aim was to lay the foundations for the search for an effective predictive model, based in future on neural networks, which would make it possible to improve the quality of work in this category

2. Methods

2.1. Literature search

The aim of this research is to take stock and assess the state of the art regarding the progress made in predicting noise disturbance due to HVAC (Heating, Ventilation and Air Conditioning) in tractor cabs from 2009 to 2023. The aim was to obtain a complete picture of the methodologies used, the psycho-acoustic parameters analyzed, and the prediction models used in this specific field of research. This preliminary phase allowed the various research steps to be carried out using a methodology that had been evaluated and validated.

The methodology used for bibliographical research consists of different steps, such as bibliographic search, selection of the most relevant papers, tabulation of the information sought and, finally, analysis of the data collected. The following chapters will provide a more in-depth analysis of the results and methodology used.

This research is based on studies that have evaluated noise and vibration in tractors and vehicles at driver's ear height, using measurements including subjective measurements to evaluate and calculate predictive models. The research, however, has clearly overlooked the specific area of interest we are seeking, namely research on HVAC-induced noise in environments such as car and tractor cabins. Studies have been carried out on noise disturbances in commercial environments and spaces, but research is still lagging in this specific area. This highlights, once again, that research needs to make progress in this branch, as noise from HVAC has a strong and negative impact on the health, comfort and psychophysical well-being of the user.

The focus of research is on articles and studies published from 2009 to 2021, as no results were found before 2009 that investigated noise prediction for HVAC systems on any kind of vehicle and car. My research has not produced specific results for predictive models of HVAC noise in the cab of a tractor, but only in passenger vehicles such as motor vehicles or large commercial spaces. For this reason, the research also considered, integrated and included in the analysis of data the results obtained from the work of Marco Favaretto [39] in his research. This has allowed the research area and resources available to be expanded. The articles examined during the research were a fundamental starting point for the construction of the methodology and study of noise from HVAC in the specific case studied, despite the lack of specific literature on this subject.

The methodology of article search is called PICO, **an acronym** whose letters mean Population, Intervention, Comparison and Outcome. For each of these categories, keywords were identified which then became the search terms. The research therefore focused on sound quality predictions in tractors and vehicles, with the specification that it was noise produced by HVAC. In addition, the research focused on the psycho-acoustic parameters evaluated and the subjective tests conducted, in terms of numbers and type. Figure 1 shows the keywords chosen for the PICO categories that produced the required results. The research was conducted on electronic resources of the Politecnico di Torino, on Scopus. After selecting the articles of greatest interest, the complete texts were searched on Google Scholar, ResearchGate and the journals to which you have access thanks to the university libraries, in case of not being found on Scopus itself.

2.1.1. Selection of papers

Once the first phase of actual search for articles was concluded, the search continued with a first phase of skimming less relevant articles by reading the title and abstract. The articles selected in this first phase were then further analyzed by reading the full text. Therefore, the most relevant papers emerged from this subgroup and were in line with the research conducted. **Table 1** shows the selection criteria for the articles examined.

The choice of articles was not only relegated to the prediction of noise from HVAC in tractor cabs, but was also extended to other areas, such as the study in vehicles, motor vehicles and environments, as the specific literature on the topic is very limited and it would not have produced constructive results.

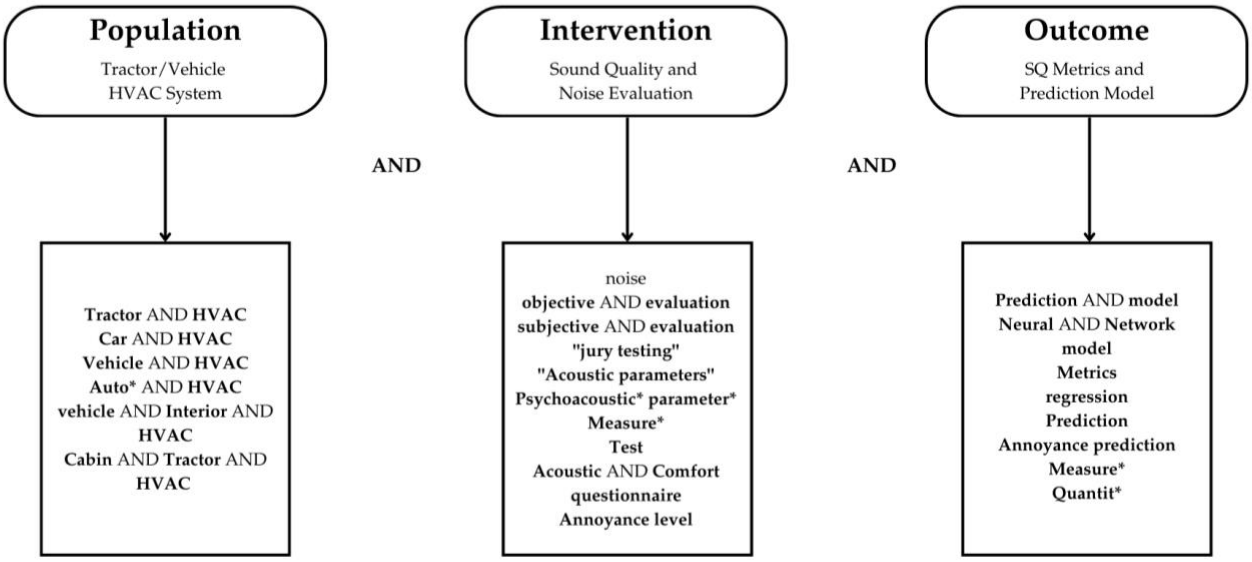
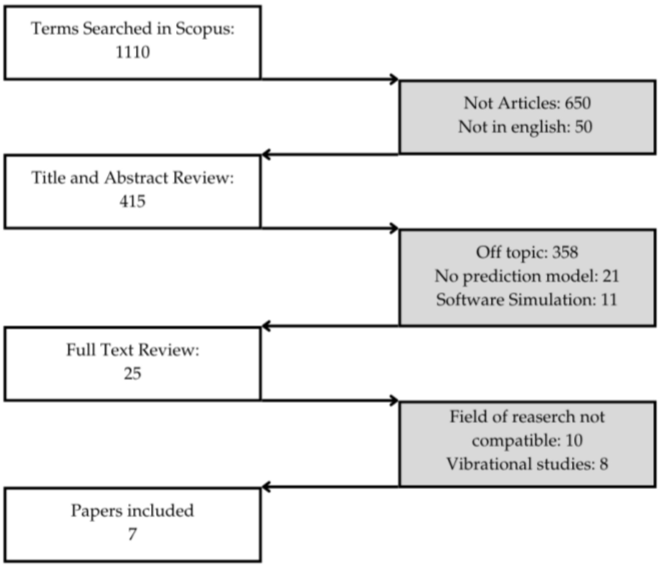


Fig.1 keywords used in the PICO search. The special character '*' determines an Inclusion of all the words with the same radix as the one in the table.

Table 1
Inclusion criteria for the selection of papers researched

Inclusion	Exclusion
• Studies related to HVAC's noise prediction • models inside tractor's cabin • Studies contain significant correlation (p < 0.05) between tractor or vehicle interior • sound and the development of a noise • prediction model • Experiments carried out in-field or in laboratory with ecological validity • Studies involving subjective testings	• All papers for which the topic was unrelated to acoustics • Review papers • Proceedings of conferences • Papers not published in English • Papers not dealing with vehicle interior noise • Papers using 3D software simulations for sound prediction • Papers for which a full text is not available • Subjective tests not involving binaural or sound listening in a controlled environment

Figure 2 represents the phases of filtering and analysis of the papers



2.1.2 Tabulation

The papers that were selected and that met the above criteria were then analyzed in order to find the key information for the research and the aforementioned were tabulated in order to have the main information clear and visible. The tabulation is presented in Table 2 and the details regarding it are given below.

2.1.2.1 Provided information

- **Vehicle:** Specifies the type of vehicle studied in the article. You can have: Fuel Vehicle (FV), Commercial Room (CR)
- **HVAC:** indicates the type of ventilation studied, such as ventilation or heating.
- **Recording:** describes how the measurements were performed, i.e. the scenario in which the vehicle is inserted during the measurement phase. The categories can be the following Different Speeds (DS), Different Roads (DR), Semi Anechoic Chamber (SAC), Idle Engine Run-up (IER), naked HVAC (N).
- **Equipment:** this category represents the number and type of microphones used for noise measurement. You can have Artificial Head (AH), Microphones (M), Binaural Microphones (M(B)), Microphone Headset (MH), Acoustic Camera (AC)
- **Mic Position:** this parameter represents the positioning of the microphone inside the vehicle during the measurement phase. You can have Operator Ear Level (OEL), Different Positions (DP) for multiple positions inside the vehicle.
- **Acoustical Parameters:** this parameter represents the acoustic characteristics analyzed in the article. Specifically: Loudness (L), Sharpness (S), Roughness (R), Fluctuation Strength (FS), Articulation Index (AI), Tonality (T), Linear Sound Pressure Level (L-SPL), A-Weighted Sound Pressure Level (A-SPL), Prominence (P), Tone to Noise Ratio (TNR), Hand Vibration (HV), Seat Vibration (SV), Waveform (WF), Mel frequency cepstral coefficients (MFCCs)
- **Subjective Evaluation Method:** this parameter explains the approaches used for subjective evaluations of sound quality, such as Rating Scale Method (RSM), Semantic Differential Method (SDM)
- **Jury:** specifies the number of people subjected to the subjective test.
- **Prediction Method:** represents the techniques and methods used to predict models. We can have Multiple Linear Regression (MLR), Software identification (SI), Neural Networks (NN), Bradley-Terry Model (BTM)

2.1.2.2 Specifications of the provided information on subjective evaluation methods

There are different methods for obtaining and collecting the subjective impressions of people who are available to participate in subjective tests. In the previous paragraph, the Rating Scale Method (RSM) is mentioned as data collection methods, which allows the subject to express a judgment on a fixed and predefined scale, which follows a criterion chosen a priori, and the Semantic Differential Method (SDM), which uses a polarized semantic scale which, consequently, allows for a more precise result.

2.1.2.3. Specifications of the provided information on prediction models

The prediction methods are varied, as are the techniques used. We have, for example, Multiple Linear Regression (MLR), Deep Belief Networks (DBN), Back Propagation Neural Network with Simulated Annealing (BPNN(SA)), Bidirectional Long Short-Term Memory with Genetic Algorithm (GA-BiLSTM), XGBoost algorithm , Back Propagation Neural Network with Genetic Algorithm (GA-BPNN), Laplacian Score Deep Belief Network (LS-DBN), Synthetic Annoyance Evaluation for ANNs (ANN-SAE), Particle Swarm Optimization BPNN (PSO-BPNN), Deep Belief Network (DBN), Mahalanobis Distance (MD), Gray System Theory (GSM), and Vehicle Noise Annoyance Neural Network Model (VNA-NNM), Convolutional Neural Network (CNN), Adaptable Learning Rate Trees CNN (ALRT-CNN), Time Frequency Images CNN (TF-CNN), Simulated Annealing and Genetic Algorithm BPNN (SAGA-BPNN).

There are many possible approaches with these types of prediction methods, such as machine learning techniques such as neural networks and all types of optimizations that implement this type of algorithm.

Linear regression statistical techniques are used to model the relationship between an independent and dependent variable with a linear equation. In this case, a linear relationship between the input variables is assumed and the combinations capable of giving the best possible linear prediction equation are calculated.

Techniques based on neural networks, in all their possible architectures, are statistical methodologies that base their operation on a mathematical model of the human brain, with interconnected nodes organized in layers. Each node analyzes an input and produces an output, which, in turn, will be the input of the next layer, until reaching the final output. The advantage of the neural network is the accuracy and the possibility of carrying out much more complex calculations and with very complex non-linear relationships between data. Therefore, the relationships between data processed by neural networks are multidimensional and with complex dependencies between input and output variables. The disadvantage is that they require a very large amount of input data.

REF	Vehicle	HVAC	Recording	Equipment	Mic Position	Acoustical Parameters	SEM	Jury	PM
[1]	FV	Yes	DS, SAC	MH	OEL	L, R, S, TNR, P, FS	RSM	27	MLR
[2]	CR	Yes	DS, SAC	AC	OEL	L, S, SPL, A-SPL	/	/	/
[3]	FV	Yes	SAC	M	OEL	/	/	/	/
[4]	FV	Yes	DS, SAC	M	OEL	L, R, S, FS	SDM	35	MLR, NN
[5]	FV	Yes	DS, SAC	/	OEL	L, R, S, SPL	RSM	30	MLR
[6]	FV	Yes	N, DS, SAC	M, AH	DP	L, R, S, SPL, A-SPL	RSM	24	BTM
[7]	FV	Yes	DR, DS, IER	MH	OEL	L, R, S, T, FS, SPL	/	/	/

Table 2. Summary of the contents collected from the papers (REF) included in the review.

2.1.3 Translation of psychoacoustic parameters for subjective tests

Psycho acoustic parameters are defined with English terminologies that have a certain depth in meaning and semantic nuances. The study is conducted in Italy and the subjects to whom the subjective test is administered are of Italian origin and language. Consequently, it was necessary to carry out a translation process of the individual parameters identified for the analysis.
The methodology followed is the one illustrated in the article by Gragnaniello et al [9] , which investigates the most appropriate and effective method for translating words or sentences into a different language.
To carry out the translations, the intervention of two native English speakers and one bilingual Italian and English speaker was necessary. The translation then took place in three phases.
In the first phase, the words to be translated are administered to a group made up of two native speakers. They carried out the translation of each parameter from English to Italian independently and autonomously.
Subsequently, the translations produced by the two subjects in the previous phase were administered to the third subject, unaware of the work carried out previously. He translated the previously produced translations from Italian into English, returning to the original English terms.
Finally, the three native and bilingual subjects had time available to discuss the correctness of the translations produced. Two acoustics experts were also present in their company, in order to provide support if necessary.
The translation results are reported in **Table 3** and then used in the subjective perception rating scales administered during the tests.

Choosen Parameters	First Phase	Second Phase	Third Phase	Results
• Fluctuation Strength: Fluctuating/Not Fluctuating • Tonality: Whistling/Not Wistling • Loudness: Loud/Not Loud • Roughness: Rough/Not Rough • Sharpness: Sharp/Not Sharp • Annoyance: Annoying/Not Annoying	• Fluctuation Strength: Fluttuante/Non fluttuante • Tonality: Fischiante/Non Fischiante • Loudness: Forte/Non Forte • Roughness: Ruvido/Non Ruvido • Sharpness: Acuto/Non Acuto • Annoyance: Fastidioso/Non Fastidioso	• Fluctuation Strength: Fluctuating/Not Fluctuating • Tonality: Whistling/Not Wistling • Loudness: Loud/Not Loud • Roughness: Rough/Not Rough • Sharpness: Sharp/Not Sharp • Annoyance: Annoying/Not Annoying	The evaluation of the translated terminology was declarated valuable and with a high level of accuracy by both the English experts and the acoustic expert	• Fluctuation Strength: Fluttuante/Non fluttuante • Tonality: Fischiante/Non Fischiante • Loudness: Forte/Non Forte • Roughness: Ruvido/Non Ruvido • Sharpness: Acuto/Non Acuto • Annoyance: Fastidioso/Non Fastidioso

Table 3. Steps and results of the translation process of psycho acoustic parameters.

2.1.4 Data analysis

The data collected through bibliographic research and state-of-the-art analysis are discussed and analyzed in the current chapter. The first section will investigate the results that emerged from the bibliographic research explored in the previous sections, the second part, however, will report the analysis of the current bibliographic research and that conducted by my colleague Marco Favaretto [39], in such a way as to have a general picture of the study of the state of Current and comprehensive art regarding the prediction of noise produced by HVAC in tractor cabs.

2.1.4.1 Current Analysis

There are no papers that specifically and directly address the problem of HVAC in means of transport such as tractors. Research has reported a high percentage of studies regarding HVAC noise in vehicles, especially in vehicles with internal combustion engines. The exception is a study conducted on noise caused by HVAC in an industrial environment. This article was kept as valid in the selection, as, apart from the subject studied, it is inherent to all the other steps and other needs that the research imposed. **Figure 3** shows a pie chart that allows you to visualize the distribution.

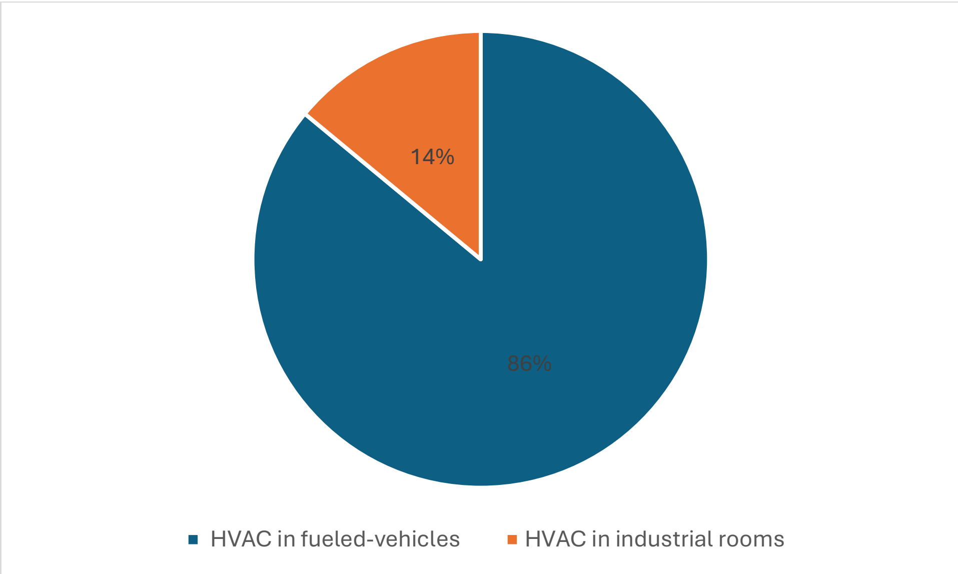


Fig. 3 percentage occurrence of vehicle types in the literature search.

The measurements carried out in internal combustion engine vehicles were mostly carried out in an acoustic chamber and with different speeds of the HVAC block tested. Only in one article were HVC levels recorded with the engine running and the car moving on different roads. In the latter, a recording was also carried out with the vehicle stationary and the engine running and one with the engine off. Furthermore, about half of the articles reported tests performed on different cars with different HVAC settings.
Some Articles did not report having conducted subjective tests or did not specify the use of prediction models. However, their value from a research point of view was considered regarding the choice of psycho-acoustic parameters and the methodology for collecting objective data.
Most sounds were detected with microphones and ambisonics microphones. Only in one case was an artificial head used for acquisition, to simulate the masking effect of the driver's head. Furthermore, all measurements in vehicles were made at the height of the operator's ears, equal to approximately 1.20m.

The psychoacoustic parameters most used in the literature are loudness, with a percentage equal to 21%, and at the same percentage score we find sharpness. A few percentage points away we find roughness, with an occurrence percentage of 17%. With the same score, 10%, we find SPL(A), SPL and fluctuation Strength. The difference between Sound Pressure Level and A-Weighted Sound Pressure Level is in the weighting curve that is applied to the value. SPL is the unweighted value, while A-SPL is the weighted value for the weighting curve corresponding to human perception. Finally, with a lower number of occurrences, we find Tone to noise ratio, Prominence and Tonality. In **Figure 4** the occurrence values are presented graphically.

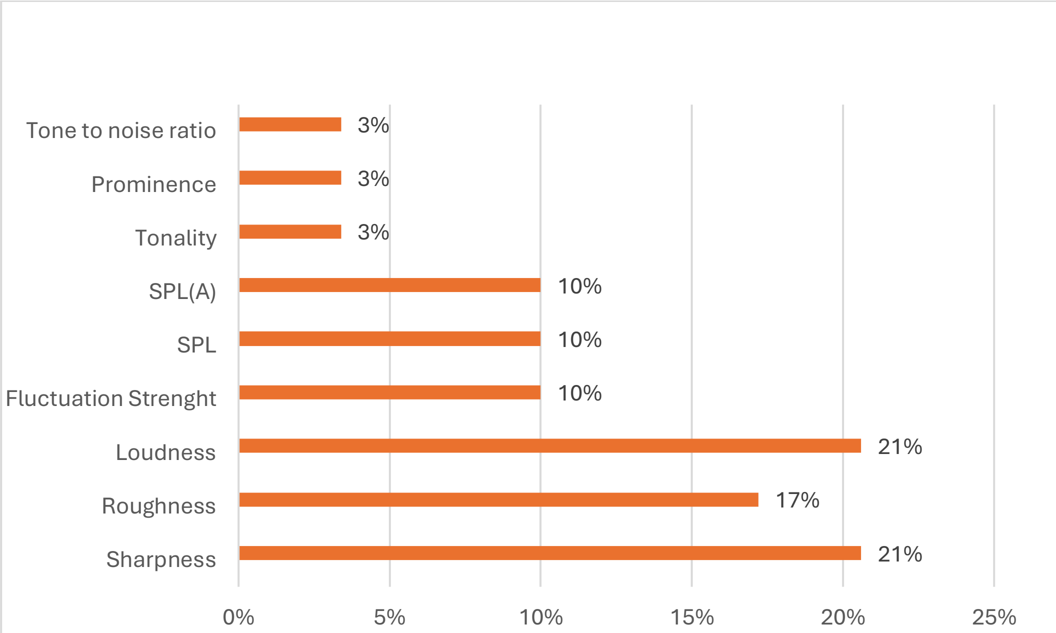


Fig 4. Number of occurrencies of Loudness, Sharpness, Roughness, Tonality SPL, A-SPL, Prominence, Tone to noise ratio, Fluctuation Strength.

As regards the subjective tests, in the selected articles an average of approximately 28 tested subjects emerges. There are many methods of evaluating subjective perception. In the selected papers, the Rating Scale Method (RSM) (28%) and the Semantic Differential Method (SDM) (14%) are recurring. The two are also used in combination (14%). As stated previously, there are articles that do not specify having taken the subjective tests; therefore, there are no data regarding all the articles chosen (43%).
Figure 5 shows the above results:

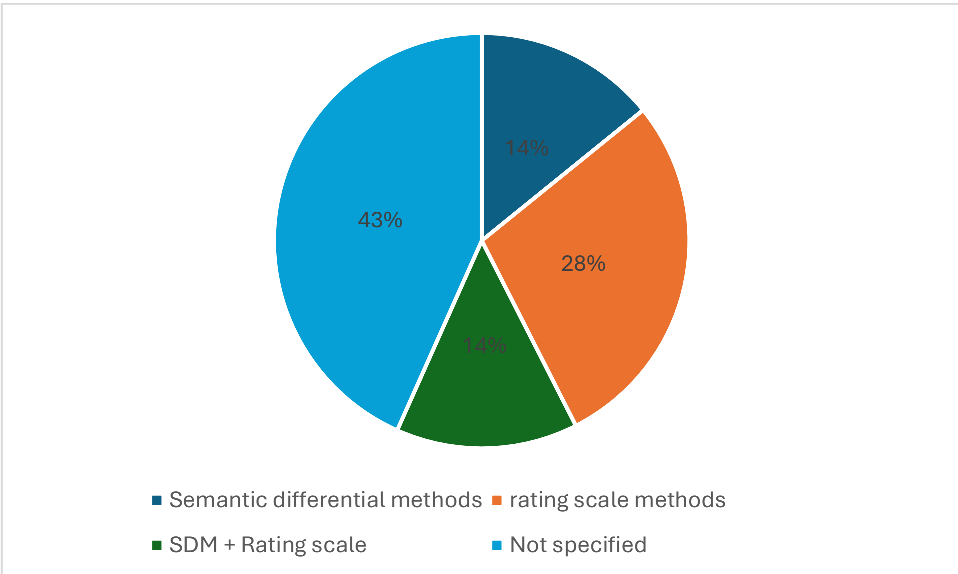


Fig. 5 occurrence, in percentage, of subjective evaluation methods used in the research, i.e., Rating Scale Method (RSM), Semantic Differential Method (SDM) and the two methods combined.

The effectiveness of the chosen predictive model is partly intrinsic to the choice of the model itself. Analyzing the literature, we can see how Multiple linear regression has a very high percentage, equal to 43%. Second to it we have the Neural Networked, with a percentage equal to 14%, therefore significantly less. Finally, with a rather high percentage, all those different methodologies used are collected, which do not fall into a precise category and are used by sporadic studies. Some of them do not even appear to be predictive methods, as they deal with identifying the specific source of air and noise in the HVAC block. **Figure 6** presents a graph that summarizes what has been said:

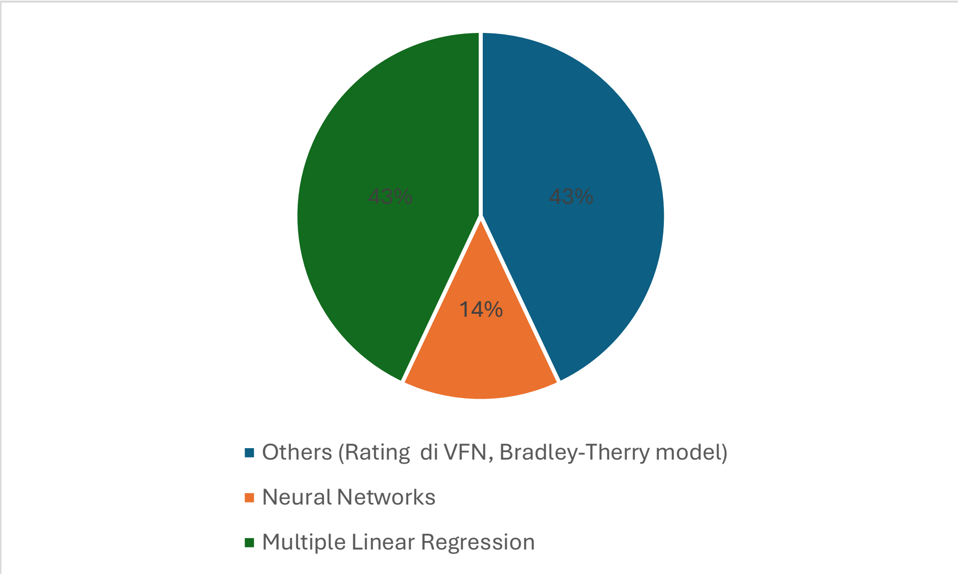


Fig.6 occurrence, in percentage, of the prediction models used in various research.

2.1.4.2 Compared Analysis

In this section, a unified analysis of the data produced by the research examined previously and that conducted by my colleague Marco Favaretto [39] will be performed, to have the broadest and most complete picture possible.

By combining the data, the population investigated becomes larger and more diverse. The heaviest percentage is once again obtained by the category of endothermic propulsion vehicles, with a value of 59%. Next come the electric vehicles. The other categories shown in **Figure 7** are not very relevant in the research.

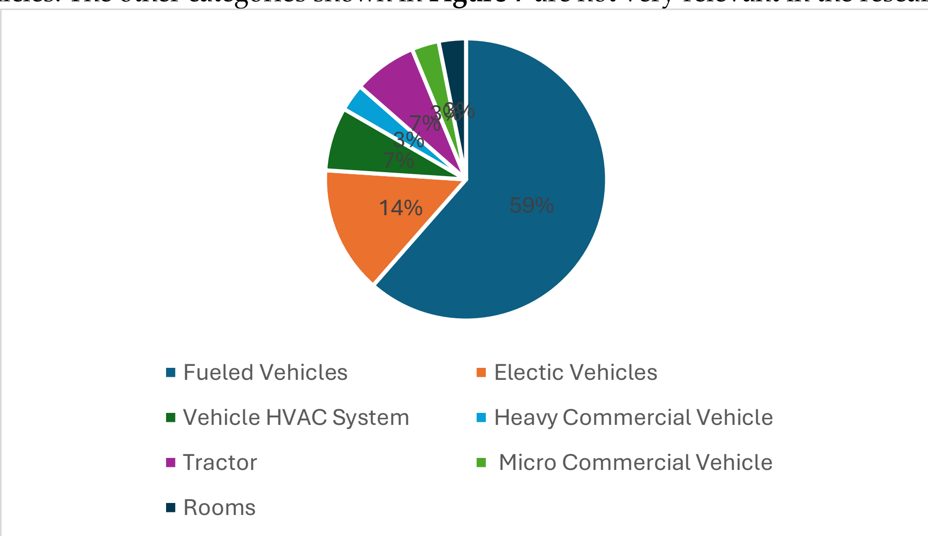


Fig.7 percentage occurrence of vehicle types in the literature search

From Doctor Favaretto's research [39] it emerged that endothermic powered vehicles are subject to driving tests on different roads with the HVAC set at different speeds. At the same time, electric vehicles, on the other hand, are tested stationary with different speeds of the HVAC block. Finally, the tractors are tested with the engine running at minimum power. These results are similar to those that emerged in the research analyzed in the previous paragraph.
The microphones used appear, in the combined analysis of the data, to be mostly those belonging to the category of binaural microphones and artificial heads.
The most used psychoacoustic parameters, in this case, are primarily Sharpness and Loudness with a percentage value of 18%. In second position we find Roughness with 16% of recurrence, followed by Fluctuation Strength (11%), SPL(A) (9%), Articulation index (8%), Tonality (7%), SPL (4%). With percentages equal to 1% we find the parameters listed in **Figure 8**.

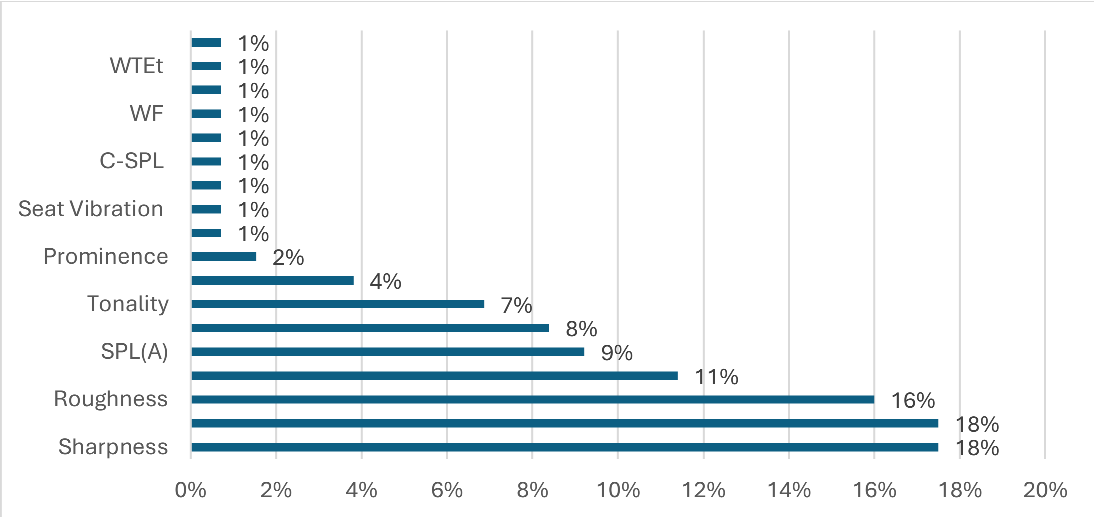


Fig.8 Percentage occurrence of the psychoacoustic parameters used in the research: i.e., Loudness (L), Sharpness (S), Roughness (R), Fluctuation Strength (FS), Articulation Index (AI), Tonality (T), Sound Pressure Level (L-SPL), A-weighted Sound Pressure Level (A-SPL), Prominence (P), Tone to Noise Ratio (TNR), Hand Vibration (HV), Seat Vibration (SV), Waveform (WF), Mel Frequency Cepstral Coefficients (MFCCs), Wavelet Transform Energy (WTEn), Wavelet Transform Entropy (WTEt);

Compared to the data from the analysis explored in the previous paragraph, we can see how a percentage superiority of the Loudness, Sharpness and Roughness parameters has emerged. The parameter referring to SPL has lost many percentage points. Overall, the results are very consistent between the two analyses.
Moving on to the subjective tests, the average of subjects tested in the articles analyzed by the colleague we have a result equal to 25 compared to 28 in the previously carried out analysis. This leads to an overall average of 26 subjects.
The prominent methodology regarding subjective tests is Rating Scale Method, with an overall percentage of 48%. Next, we find Semantic differential method. The other values shown in the graph in **Figure 9** are of little relevance to the research given their recurrence. Again, the results are consistent with the research reviewed previously.

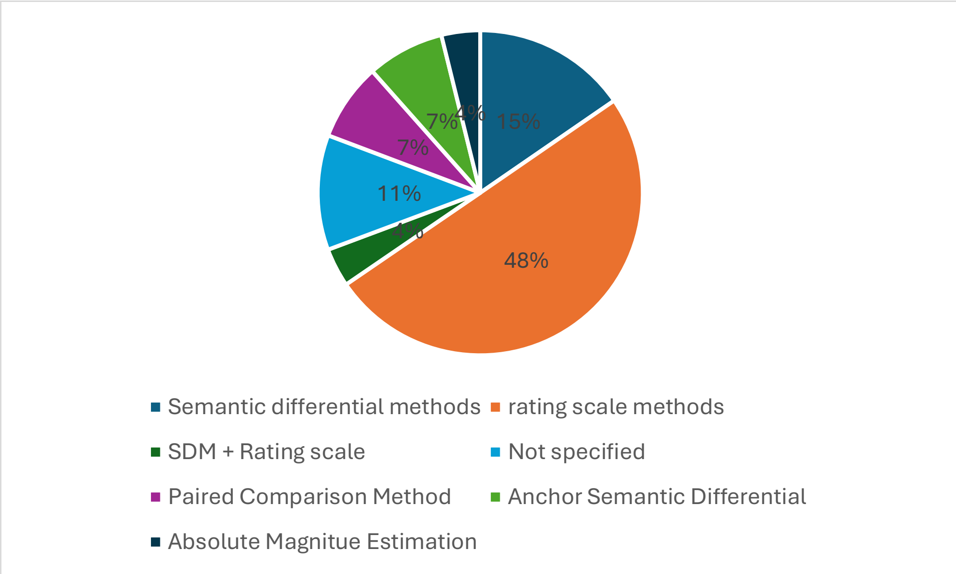


Fig. 9 occurrence, in percentage, of subjective evaluation methods used in the research, i.e., Paired Comparison Method (PCM), Rating Scale Method (RSM), Semantic Differential Method (SDM), Absolute Magnitude Estimation (AME), Semantic Differential with Anchor Stimulus (ASD), SDM + RSM.

As regards the choice of the predictive model, the results were reversed compared to the research investigated previously. In fact, the use of Neural networks occurs with a recurrence rate of 55%. Multiple linear regression falls to second place, with a percentage score of 22%. Other methods that emerged are shown in **Figure 10**, but with very low percentages, making them very irrelevant.

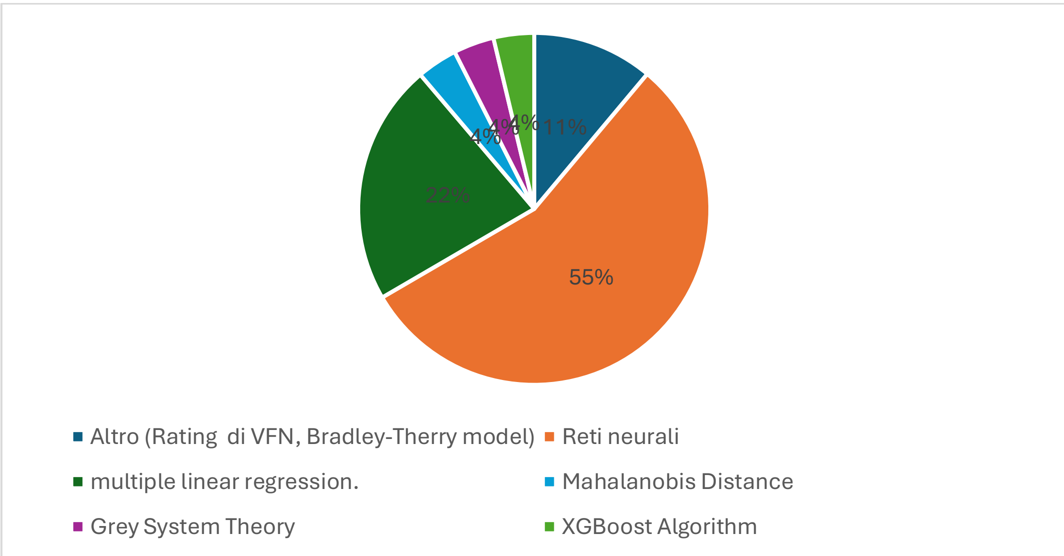


Fig.10 occurrence, in percentage, of the prediction models used in various research.

From Marco Favaretto's research [39] an interesting aspect emerged regarding the correlation between parameters regarding their use in prediction methods. It has been shown that the most influential parameters regarding prediction are Loudness, Sharpness and Roughness, as they present a high general correlation.

2.2 Signal Acquisition

The noise acquisitions were carried out on 8 July 2024, inside a Lamborghini brand tractor, made available by the company DENSO. It is precisely in their headquarters located in Poirino, a municipality not far from Turin, that we carried out the measurements.
The measurements were taken outside the shed where the tests were carried out, in a quiet and little-used area of the car park, overlooking a wide but little-trafficked road. The measurements were carried out outside as there was a fan in the internal area which needed to remain on as it was useful for other endurance tests in progress.
The HVAC block of the tractor on which the measurements were made was located on the roof of the tractor. The tractor remained with the engine off for the entire duration of the test and the HVAC was driven by an external battery which allowed the setting to be chosen.
There are eight vents on the vehicle and for all tests they were kept open in the same way. They are located on the upper part of the tractor, arranged in a circular manner.
The HVAC has the possibility of operating in two modes: "DEF", abbreviation which indicates the De-Frost function, and "VENT", which indicates the Fan mode. The data reported were acquired in "VENT" operating mode.

Fig.11 Position of the tractor During the measurements.

Fig.12 External battery used for Powering the HVAC.

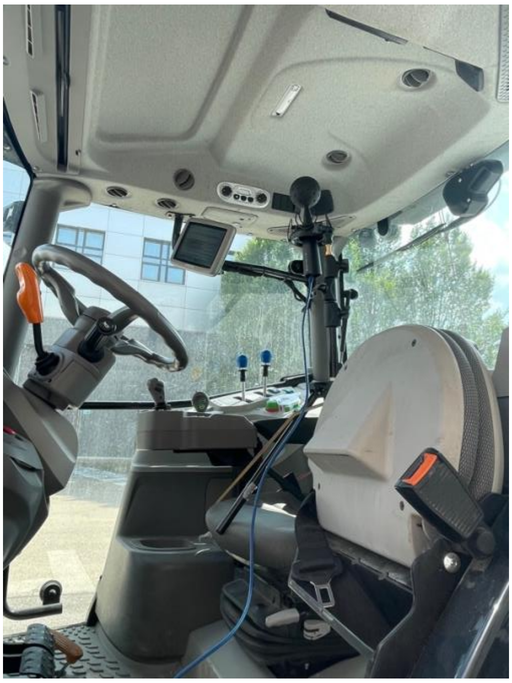
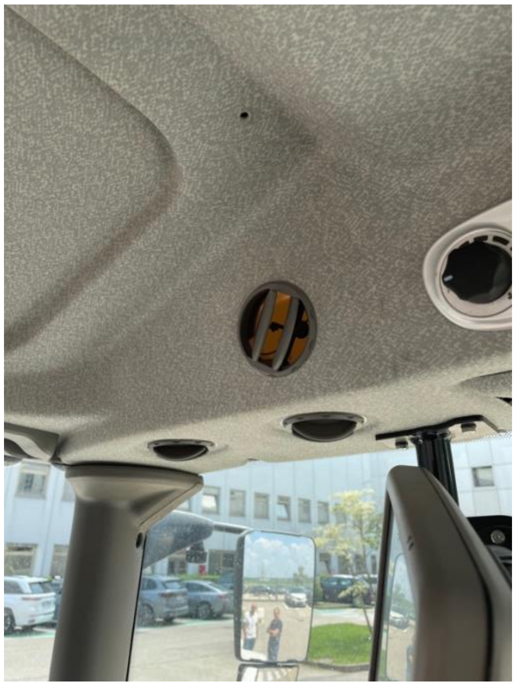
 

Fig.13, 14 Position of the air vents in the tractor cab.

The measurements were taken at different HVAC speeds. Ventilation had a range from 1 to 12, therefore significant speeds were chosen at which to carry out the measurements.
The speed settings to be recorded were managed externally by the battery shown in **Figure 12**. The device allows you to adjust the voltage level with which to power the HVAC block and, consequently, adjust its speed. Furthermore, a measurement of the levels was carried out with the HVAC off, in order to have a comparison value. Table 4 lists the recorded speeds and their acronyms, which will be used in the analysis.

Mesurements	Abbreviations
• Background NOISE	• BG
• Velocity 4	• V4, Vel4
• Velocity 8	• V8, Vel8
• Velocity 12	• V12, Vel12

Table 4 Velocities measured and their abbreviation

The measurements chosen are, therefore, speed 4, 8 and 12. The choice fell on these specifications, as they sufficiently covered the speed spectrum available, whether to have measurements that were little different from each other.

2.2.1 Microphones

The microphones used for the measurements were positioned at a height of 1.20m from the floor of the tractor cockpit. The accepted tolerance for microphone placement is +0.05m. In this way we ensure that the results are at least consistent with each other, which is fundamental for comparing psychoacoustic measures.
The duration of the recordings made is equal to 60 seconds for speeds 4, 8 and 12. The recording relating to the background lasts 5 minutes. Three different measurements were performed for each speed and for each microphone, in order to have at least one interference-free recording for each microphone.
The microphones used are the Zylia ZM-1, NTi M2230 and the artificial head produced by HeadAcoustics model HSUIII.3. This latter microphone did not require calibration, as the parent company provided the specifications of the current calibration; therefore, the factory one was maintained as it was known. The NTi has been calibrated with its supplied calibration device, which outputs a frequency at 1kHz. The Zylia ZM-1 underwent a calibration process following the measurements. The recordings made with the ZM-1 were played through headphones to the HeadAcoustics artificial head used for the measurements and the gain was manually modified in order to reach a level equal to that recorded by the Head3.III, whose calibration we knew.

2.2.1.1 Zylia ZM-1

- Positioning: operator ear level
- Frequency range: 20Hz - 20kHz
- Type of recording: ambisonics audio

The Zylia ZM-1 is a sphere-shaped ambisonic microphone that features 19 acquisition channels evenly distributed across its surface. The 19 capsules use MEMS technology. The positioning of the capsules is such that the microphone is able to capture the sound coming from every direction.
The recordings acquired with this microphone were used to conduct the subjective tests. The tests will be carried out in the anechoic chamber available at the Polytechnic of Turin. The subjects will have access to a virtual reality viewer, in which they will be shown a 360° video of the inside of the tractor cab, and a pair of Sennheiser 650HD headphones, where they will be able to hear the audio recorded on the spatialized tractor.
To carry out the measurement, the microphone was connected to a laptop and the recording was managed with the support of the Bidule software.

Fig.15 Zylia ZM-1 microphone
In the tractor during measurements

Fig.16 Zylia ZM-1 microphone

The MEMS technology used by the Zylia ZM-1 allows for a rather flat normalized impulse response between 50 Hz and 15 kHz, with a poor response at very low and very high frequencies, as shown in **Figure 17**. This is not optimal given the type of noise detected, which, as we will see in the analysis, has a strong low-frequency component.

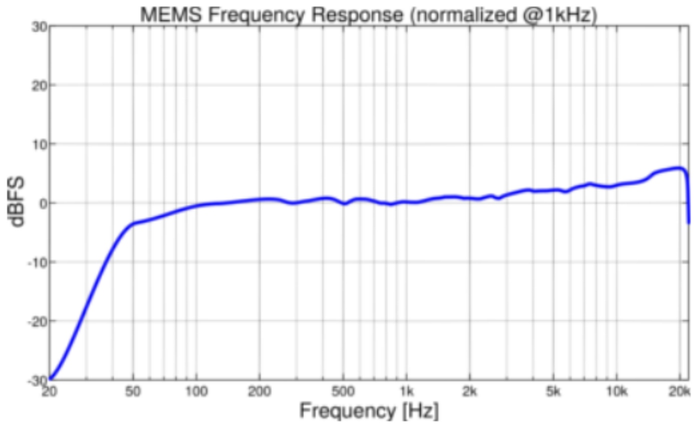


Fig. 17 Zylia ZM-1 impulse response normalized

2.2.1.2 NTi M2230

- Sensitivity typical @1kHz: 42mV/Pa
- Measured Quantity: pressure
- Frequency range: 5Hz - 20kHz
- Positioning: operator ear level

The NTi microphone is a type of omnidirectional microphone, whose acquisitions were used for the analysis and extrapolation of the objective data analyzed later. The microphone has a preamplifier integrated into the body and is powered by a 28V phantom. For acquisition, the microphone connects to XL2 Audio and Acoustic Analyze, which allows you to set a timer for recording, evaluate levels, generate reports at the end of recording and allow you to carry out calibration.

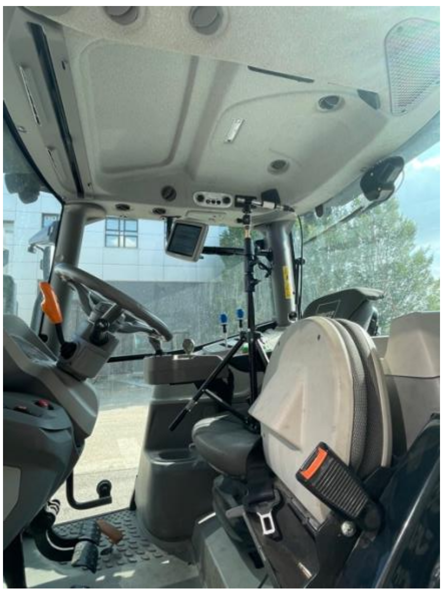


Fig.18 NTi microphone
In the tractor during measurements

Fig.19 NTi XL2 acoustic analyzer

Figure 20 shows the impulse response of the NTi omnidirectional microphone, which demonstrates a very flat response for both very low and very high frequencies, even if in the latter it has some more irregularities.

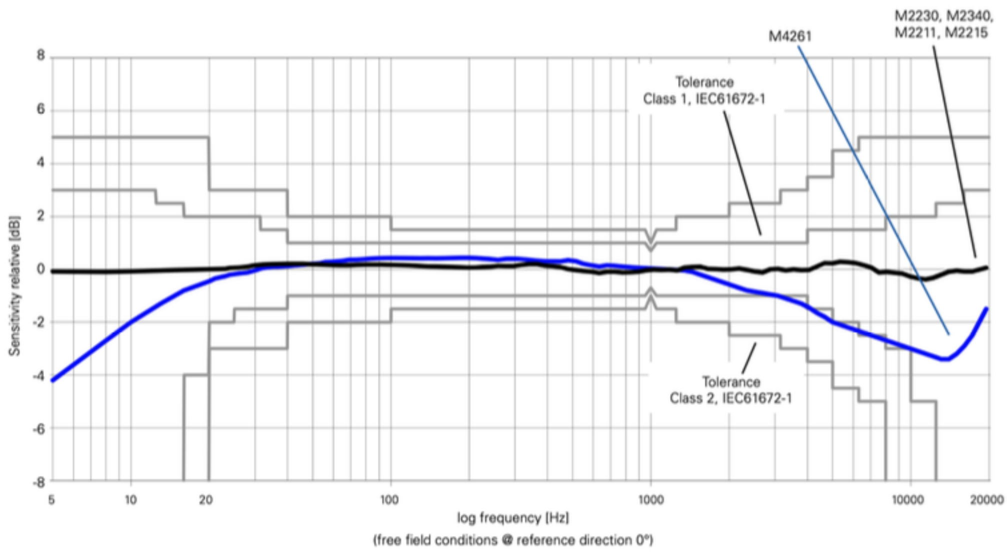


Fig.20 Impulse response of NTi microphone in free field measurements (black line)

2.2.1.3 HSU III.3

- Low inherent noise of 6.5 dB(A)
- Dynamic range of 106.5dB(A)
- Positioning: operator ear level
- Frequency range: 6Hz – 20kHz

The Head Acoustics HSU3.III artificial head is an artificial head that also features a neck and torso and has two ½" sized microphones built into the ears. The microphones are of the high-quality condenser type. The main feature is the very low background noise that the head guarantees, equal to 6.5dB(A).
To carry out the measurement, the artificial head was placed on a stand and positioned, like the other microphones, at ear height of the driver. Furthermore, for recording, it is necessary to connect the artificial head to an audio interface which, in turn, is connected to the computer, where, using the Head Acoustics home software "ArtemisSuite", recording can be carried out.

Fig.21 HSUIII.3 in the tractor cabin during measurements

Fig.22 HSU 3.III

Fig.24 Audio interface and battery used

For this study, we had the possibility to use *B&K 4101*, the binaural microphone with artificial head and torso used by Marco Favaretto in his research [39]. The decision of using HeadAcoustics HSU 3.III was made because of the characteristics of the microphone itself, such as having low-noise microphone, a 32bit dynamic range with double ADC converter, maximum sampling rate of 204.8kHz and microphones equalization with digital filters directly on the instrument. Those characteristics allow to record the whole dynamic range, from hearing threshold to pain threshold. Therefore, with this instrument we made more precise measurements.

2.2.2 Psychoacoustic parameters

Psycho-acoustic parameters are objective values, derived from formulas, which allow us to have a numerical value capable of best representing and approximating the human perception of some sound characteristics. This makes them fundamental, as sound and its characteristics are particularly relevant in relation to man and human perception. The parameters are derived from specific acoustic characteristics of the sound, but go beyond mere objective perception, providing much more useful and relevant information.
In the following paragraphs, the parameters used in the research will be investigated and explored in depth, with particular attention to the definition, applications and just noticeable difference. The information regarding the psychoacoustic parameters was acquired and reported from the manual "Psychoacoustics Facts and Models", Hugo Fastl, Eberhard Zwicker [35] and the references were exploited those used by Marco Favaretto in his Master Thesis work [39], in such a way as to have a background and a framework as united as possible, with the aim of comparing the data collected and comparing the methodologies.
In addition to those parameters, we choose to analyze some descriptors that are based on the characteristics of the spectrum. This decision was made because of the results shown in the article written by Antti Kuusinen [8], that reports that these parameters based on the spectrum are accurate for annoyance prediction.

2.2.2.1 Linear Equivalent Sound Pressure Level

The LZeq (Equivalent Continuous Sound Pressure Level) express the consistent sound pressure level that, within a specified timeframe, encompasses the equivalent total energy of the fluctuating noise. Therefore, LZeq essentially represents the Root Mean Square (RMS) sound level, wherein the duration of measurement functions as the averaging period. This definition highlights LZeq's role as a standardized metric for analyzing the overall intensity of sound experienced over a specific duration. Replacing the Linear or Flat, the 'Z' leKer is defined as being a flat frequency response of 8Hz to 20kHz ±1.5dB.

2.2.2.2 A-weighted Sound Pressure Level

The LAeq (A-weighted Equivalent Continuous Sound Pressure Level) is the equivalent continuous sound pressure level with a standard weighting of audible frequencies ('A' weighting) that reflects human ear's response to noise. In fact, this weighting gives more the impression of the human perception of the sound. If the measurements are made with this frequency weighting, it will be displayed as dB(A) or dBA, in order for it to be clear.

2.2.2.3 Loudness

Loudness, as defined by Fastl & Zwicker [35], refers to the perceived intensity of a sound. This parameter depends on both the sound's acoustic characteristics and the listening environment, as experienced by individuals with normal hearing. While primarily influenced by the sound pressure level, loudness is also affected by factors such as frequency, waveform, bandwidth, and sound duration. Its unit of measurement is the *sone*, where one sone corresponds to the loudness of a sound with a level of 40 phons. A sound perceived as twice as loud as another corresponds to double the number of sones. In this study, according to Favaretto's research [39], loudness was calculated using Zwicker's method (ISO 532:1975, Method B), which employs diffuse field equalization. This approach was selected because, according to ISO 532-1:2017 (Annex D, page 52), free-field equalization is recommended only when a single acoustic source is present in a free field, positioned more than 1.5 meters from the head and torso simulator, with both azimuth and elevation angles at 0°. In the tractor cabin, multiple noise sources are present, including the engine, reflections from the cabin walls, and the various fans of the HVAC system. Consequently, loudness, as defined by ISO 532B, can be calculated as follows:

$$N' = 0.08\left(\frac{E_{TQ}}{E_0}\right)^{0.23}\left[\left(0.5 + 0.5\frac{E}{E_{TQ}}\right)^{0.23} - 1\right] \quad (1)$$

where E represents the sound excitation, $E0$ denotes the excitation at benchmark sound intensity, and ETQ refers to the excitation at the absolute listening threshold. The distribution of loudness across the critical bands, expressed with the unit "sone/Bark", is exhibited by the specific loudness N' (1). Integrating the critical band rate on the N' value, the total loudness N (2) results:

$$N = \int_0^{24Bark} N'(z)dz \ (sone) \quad (2)$$

Pedrielli & Cadretti [36] highlight that the just noticeable difference (JND) in loudness becomes more significant as the overall sound pressure level (SPL) of the signal increases. Given that these are ground vehicles, it can be assumed that the maximum SPL is approximately 80 dB, making the smallest perceivable difference in loudness equal to 0.8 sone. In this study, conducted on a tractor, a JND value of 0.8 sone was adopted.

When measurements are performed using an artificial head, it is preferable to report the values for both channels. A representative single-digit value can then be determined by taking either the maximum or the average of the two channels.

2.2.2.4 Roughness

Roughness is a perceptual sensation caused by rapid changes in sound due to modulation frequencies ranging between approximately 15 and 300 Hz. Exact periodic modulation is not necessary; however, the modulating function's frequency spectrum must fall within this range to create the sensation of roughness. For instance, a 1 kHz tone at 60 dB, modulated in amplitude at 100% with a modulation frequency of 70 Hz, results in a roughness value of 1 asper. In amplitude modulation, the key factors influencing roughness are the modulation depth and modulation frequency. For frequency modulation, the relevant parameters are the frequency modulation index and modulation frequency.

$$R = 0.3 f_{mod} \int_0^{24\ Bark} 20 log \left[\frac{N'_{max}(z)}{N'_{min}(z)}\right] dz\ (asper) \quad (3)$$

The studies of Fastl and Zwicker [35] have revealed that in the case of amplitude-modulated pure tones an increment of roughness becomes perceptible if the degree of modulation is increased by 10%, the corresponding increment in roughness is 17%.

2.2.2.5 Sharpness

The numerical measure of sensation based on the number of high-frequency components in a sound is represented by the psychoacoustics metric called Sharpness. To correctly define and understand the parameter under consideration, it is necessary to specify the unit of measurement used to express it, namely 'acum.' The sharpness of a broad-band noise, which has a center frequency of 1 kHz, a width of one critical band, and a level of 60 dB, is the definition of 1 acum. The scale of measurement is linear, so if the frequency content is doubled, the sharpness double as well. According to Fastl & Zwicker [35], sharpness increases for a level increment from 30 to 90 dB by a factor of two. This means that the dependence on level can be ignored as a first approximation, especially if the level differences are not very large. The most important parameters influencing sharpness are the overall spectral content and the center frequency of narrow band sounds and is not dependent on loudness level or the detailed spectral content of the sound. It is calculated as a weighted sum of the specific loudness levels contributing to the overall loudness. Several weighting functions exist (by Aures, Fastl, Von Bismarck) but they all increase toward the highest frequency bands to model the fact that the hearing system is more sensitive to high-frequency components level.
The calculation of sharpness, as outlined in DIN 45692, relies on an initial loudness calculation performed using the Zwicker method:

$$g(z) = \begin{cases} 1, & z \leq 15.8\ Bark \\ 0.85 + 0.15 e^{0.42(z-15.8)}, & z > 15.8\ Bark \end{cases} \quad (4)$$

$$S = 0.11 \frac{\int_0^{24\ Bark} N' g(z) z\, dz}{\int_0^{24\ Bark} N'\, dz}\ (acum) \quad (5)$$

Here, S (5) represents the sharpness to be calculated, and the denominator provides the total loudness N (2). The upper integral functions similarly to the first moment of specific loudness over the critical-band rate, but includes an additional factor, $g(z)$ (4), which depends on the critical-band rate and increases beyond 16 Bark. The just noticeable difference in sharpness, defined as the smallest variation detectable by at least 75% of the test subjects, is estimated to be 0.04 acum.

2.2.2.6 Fluctuation Strength

The perception of fluctuation strength occurs at low modulation frequencies, up to approximately 20 Hz, while at higher modulation frequencies, the sensation of roughness is experienced. A fixed point is therefore defined for a 60-dB, 1-kHz tone 100% amplitude- modulated at 4Hz, as producing 1 vacil.
Fluctuation strength can be described with the following expression:

$$F = \frac{0.008 \int_0^{24\ \mathrm{Bark}} \Delta L_E(z) dz}{(f_{\mathrm{mod}}/4\mathrm{Hz}) + (4\mathrm{Hz}/f_{\mathrm{mod}})}\ (vacil) \quad (6)$$

$$\Delta L_E(z) = 20log_{10}\left[\frac{N'_{max}(z)}{N'_{min}(z)}\right] \quad (7)$$

where f is the modulation frequency, and ΔL (z) is the variation of the sound signal. 012 "

2.2.2.7 Tonality

Tonality is a psychoacoustic parameter that describes how the tonal quality or character of a sound is perceived. It relates to the timbre of a sound and how closely it resembles a specific musical note. Tonality is strongly connected to the perception of a sound's frequency and harmonics, which influences its perceived "musicality." Although tonality doesn't have a specific physical unit, it is often expressed as a percentage (%) or through qualitative scales (e.g., tonal vs. atonal) to indicate how much a sound is perceived as having a clear pitch. The perception of pitch plays a role in how sounds are recognized and differentiated from one another. Sounds with a strong pitch are typically associated with well-defined musical notes, while those with a weaker pitch might be perceived as noise or less distinct. Sounds with a more defined pitch tend to stand out more in audio recordings and are often considered more pleasant or musical. Tonality is shaped by the fundamental frequency of a sound, its harmonics, and spectral content. The fundamental frequency largely determines the pitch of a sound. When a sound has a clear fundamental frequency, it produces a more distinct tone. The presence of well-defined and proportional harmonics relative to the fundamental frequency enhances a sound's tonal clarity. On the other hand, sounds with complex spectral content tend to have a less defined tone, while simpler spectral structures result in a "purer" tone. Tonality is also crucial in psychoacoustics, where it helps in understanding how the human brain perceives and identifies sounds. A classic method for measuring tonality, as proposed by Zwicker, relies on analyzing the sound spectrum. Tonality can be calculated as:

$$Tonality = \sum wi\ x\ HarmonicNumber(i) \quad (8)$$

we represent the weight attributed to the i-th harmonic and HarmonicNumber(i) the harmonic number of i.× This formula considers the harmonics present in the sound and how each harmonic contributes to the overall perception of pitch.

2.2.2.8 Articulation index

The Articulation Index (AI) serves as a measure indicating how background noise levels can impact human speech comprehension, ranging from 0% (no speech understood) to 100% (complete speech clarity). Initially developed for assessing speech privacy and communication system effectiveness, the AI metric has expanded its applications. Today, it's utilized to evaluate factors like vehicle interior noise, the sound levels of household appliances, and other areas where speech intelligibility is crucial.

$$AI = \frac{M_1}{M_{total}} \times 100\ \% \quad (9)$$

Here, M_1 represents the number of sound units that were listened to, while M_{total} denotes the total number of sound units.

2.2.2.9 Speech interference Level

Leo Beranek (1947) introduced the Speech Interference Level (SIL) to assess the impact of noise on speech communication within passenger aircraft. SIL is calculated as the average of sound pressure levels measured across octave bands (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz) and is expressed in dB. This metric provides a single-number rating and serves as a practical way to evaluate how noise interferes with speech communication, both indoors and outdoors.

2.2.2.10 Noise Criteria (NC)

Noise Criteria is a parameter based on the spectrum of the sound that is used for measure and evaluate the background noise level in indoor environment, typically related to HVAC (Heating, Ventilation, and Air Conditioning) systems. It helps assess how intrusive the background noise is to human comfort and communication. Noise Criteria uses a set of reference curves (NC curves) that represent different levels of acceptable background noise. These curves correspond to sound pressure levels (in decibels) across different frequency bands (typically from 63 Hz to 8000 Hz). For this study, the NC curves used to calculate the parameter are from the ANSI-S12-2-2008 standard [41].

	Octave-band-center frequency in Hz									
	16	31.5	63	125	250	500	1000	2000	4000	8000
NC-70	90	90	84	79	75	72	71	70	68	68
NC-65	90	88	80	75	71	68	65	64	63	62
NC-60	90	85	77	71	66	63	60	59	58	57
NC-55	89	82	74	67	62	58	56	54	53	52
NC-50	87	79	71	64	58	54	51	49	48	47
NC-45	85	76	67	60	54	49	46	44	43	42
NC-40	84	74	64	56	50	44	41	39	38	37
NC-35	82	71	60	52	45	40	36	34	33	32
NC-30	81	68	57	48	41	35	32	29	28	27
NC-25	80	65	54	44	37	31	27	24	22	22
NC-20	79	63	50	40	33	26	22	20	17	16
NC-15	78	61	47	36	28	22	18	14	12	11

Table 5 – NC Curves from ANSI-S12-2-2008 standard.

The parameter is calculated with the tangent method: the spectrum of the sound is plotted with the curves and the NC value of the sound analyzed is the nearest NC curves that is not touch by the sound spectrum.

2.2.2.11 Noise Rating (NR)

Noise Rating is a spectral parameter similar to the Noise Criteria. While the Noise Criteria is widely used in North America, Noise Rating is more diffused in Europe and Asia. The aim of the parameter is similar, it is a standard used to determine the acceptability of background noise in various indoor environments, such as offices, industrial facilities, public spaces, and HVAC systems, by comparing measured sound pressure levels to a set of established NR curves. The reference curves used in this analysis are the standard ISO curves [41].

Noise Rating (NR) curve	Maximum Sound Pressure Level (dB)								
	Octave band mid-frequency (Hz)								
	31.5	62.5	125	250	500	1000	2000	4000	8000
NR 0	55	36	22	12	5	0	-4	-6	-8
NR 10	62	43	31	21	15	10	7	4	2
NR 20	69	51	39	31	24	20	17	14	13
NR 30	76	59	48	40	34	30	27	25	23
NR 40	83	67	57	49	44	40	37	35	33
NR 50	89	75	66	59	54	50	47	45	44
NR 60	96	83	74	68	63	60	57	55	54
NR 70	103	91	83	77	73	70	68	66	64
NR 80	110	99	92	86	83	80	78	76	74
NR 90	117	107	100	96	93	90	88	86	85
NR 100	124	115	109	105	102	100	98	96	95
NR 110	130	122	118	114	112	110	108	107	105
NR 120	137	130	126	124	122	120	118	117	116
NR 130	144	138	135	133	131	130	128	127	126

Table 6 – NR Curves from ISO Standard

Noise Rating is also evaluated with the tangent method.

2.2.2.12 Preferred Noise Criterion (PNC)

Preferred Noise Criterion is a spectrum-based parameter. It is an improvement on the traditional Noise Criteria (NC) system, designed to address some of its limitations, especially in terms of subjective perception of noise quality and aims to provide a better representation of how humans perceive background noise, particularly in more sensitive environments like auditoriums, concert halls, and other spaces where sound quality is critical. Therefore, Preferred Noise Criterion is a system of curves used to measure and evaluate the acceptability of background noise in indoor environments, with a greater focus on how humans subjectively experience different noise frequencies.
The set of curves used as a reference for the analysis are the ones proposed by Leo Beranek in 1971 [40].

Preferred noise criterion curves	31.5 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
PNC-15	58	43	35	28	21	15	10	8	8
PNC-20	59	46	39	32	26	20	15	13	13
PNC-25	60	49	43	37	31	25	20	18	18
PNC-30	61	52	46	41	35	30	25	23	23
PNC-35	62	55	50	45	40	35	30	28	28
PNC-40	64	59	54	50	45	40	35	33	33
PNC-45	67	63	58	54	50	45	41	38	38
PNC-50	70	66	62	58	54	50	46	43	43
PNC-55	73	70	66	62	59	55	51	48	48
PNC-60	76	73	69	66	63	59	56	53	53
PNC-65	79	76	73	70	67	64	61	58	58

Table 6 – Preferred Noise Criterion Curves by L. Beranek.

Preferred Noise Criterion is also evaluated with the tangent method.

2.2.2.9 Room Noise Criterion (RNC)

Room Noise Criterion is a spectrum-based parameter that aims to improve the traditional Noise Criteria (NC). It is specifically developed to address the limitations of NC in rooms with non-steady or fluctuating noise sources, such as those caused by modern HVAC systems or mechanical equipment. It helps better evaluate and control the impact of intermittent or time-varying noise in indoor environments. So, it is a standard used to measure and assess the background noise in spaces where the noise levels are not constant but fluctuate over time. It accounts for the cyclic or intermittent noise commonly associated with mechanical systems like variable air volume (VAV) systems.
The RNC Curves used for this study are from the ANSI-S12-2-2008 standard [41].

	Octave-band center frequency, Hz									
RNC-curve	16	31.5	63	125	250	500	1000	2000	4000	8000
RNC-70	101	96	91	86	81	76	72	68	64	60
RNC-65	96	91	86	81	76	71	67	63	59	55
RNC-60	91	86	81	76	71	66	62	58	54	50
RNC-55	86	81	76	71	66	61	57	53	49	45
RNC-50	81	76	71	66	61	56	52	48	44	40
RNC-45	79	74	68	62	56	51	47	43	39	35
RNC-40	78	71	64	58	51	46	42	38	34	30
RNC-35	76	69	61	54	46	41	37	33	29	25
RNC-30	74	66	58	49	41	36	32	28	24	20
RNC-25	73	64	54	45	36	31	27	23	19	15
RNC-20	71	61	51	41	31	26	22	18	14	10
RNC-15	69	59	48	37	26	21	17	13	9	5
RNC-10	68	56	44	33	21	16	12	8	4	0

Table 7 - Room Noise Criterion Curves standard ANSI-S12-2-2008.

Room Noise Criterion is also evaluated with the tangent method.

2.3 Subjective Assessment

Subjective tests are one of the fundamental arts of the methodology, as they allow us to verify the accuracy of the psycho-acoustic parameters and to grasp the impressions that the noise generated by the HVAC generates on a sample of subjects. This is particularly relevant since what the study is interested in is precisely how potential consumers perceive the noise they would be subjected to.
The subjective tests were carried out in three ways, with different methods of use, in order also to test which method is the most accurate and which produces the most reliable results. The choice to have multiple playback methodologies has the objective of evaluating the accuracy of reproduction, as the device influences the result, due to its characteristics.
The subjective tests, therefore, involve a pool of different people for each test who are made to listen to 10 seconds of sound coming from the HVAC of the tractor analyzed.
The subjects tested, for each subjective test, are 15, for a total of 60 subjects, aged between 23 and 55, with a small percentage of them having hearing defects. The recordings mentioned in **Table 4** were cut to 10 seconds, taking care to keep the audio sections unaffected by noise.

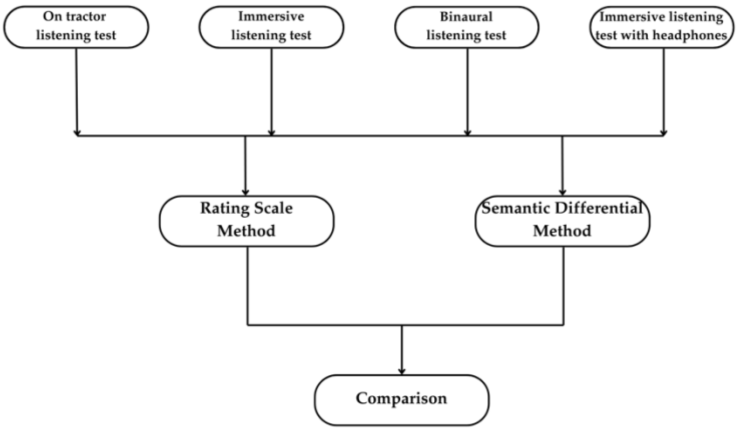


Fig. 25 Subjective Assessment Pipeline

The tests carried out are, as shown in **Figure 25**, three, with different methodologies. The first test carried out involved fifteen participants on the day of the objective measurements. The subjects took turns getting on the tractor and listened to the noise produced by the HVAC live. The second method involves the use of the Audio space lab in Politecnico of Turin. The laboratory is third order ambisonics and allows the subject to experience the sound from the tractor with 16 loudspeakers. In order to make the experience more immersive, a virtual reality headset that reproduce a 360° video recorded in the tractor is given to the subject. The third test conducted involved fifteen subjects in the anechoic chamber of the Polytechnic of Turin. The playback method involved the use of Sennheiser 650HD headphones and the Artemis Suite software, which allowed the binaural playback of the recordings and the collection of the subjects' responses. The last method reported in **Fig.25** is immersive reproduction which, thanks to the support of the virtual reality viewer and the spatialized audio via the Reaper software and the FB360 plugin, allowed the tractor environment to be virtually recreated. In this case, user responses were recorded on an Excel sheet.
All subjects in all types of tests were subjected to two types of measurement scales, namely Semantic Differential Method and Rating Scale Method, in accordance with the results that emerged from the bibliographic research and the study of the literature. In particular, the references used to create the scales can be found in Leite's papers regarding Semantic Differential Method and in Chen's article for Rating Scale Method.
In order to minimize the statistical variability introduced by the large number of different subjects tested, 5 subjects were selected to whom all three types of tests in the laboratory were administered.

2.3.1 Rating Methods

Rating methods were applied equally to all subjects of all three types, on different software and graphics, but with the same parameters and methodology. This was done to have maximum compatibility between the tests performed and to carry out an accurate comparison.
For the tests carried out at the Polytechnic of Turin, therefore the immersive and binaural ones, the subjects were offered an example audio regarding the more complex psycho-acoustic parameters, such as sharpness and roughness. This is to have compatibility with the tests carried out in Marco Favaretto's research [39], with which the results will be compared, and to improve the accuracy of the test itself, as suggested in Norm's research. For the test carried out on the tractor this preliminary step was not carried out, as the tested subjects were partly already familiar with these concepts and due to a lack of equipment.
The psycho-acoustic parameters present in the rating scales were translated into Italian as the subjects tested were entirely Italian following the methodology reported in the previous paragraphs and in Table 3.

2.3.1.1 Rating Scale Method

The 1-10 rating scale concerns the subjective annoyance parameter. The scale has five categories, which are expressed by numerical ranges. They range from 'Little Annoyance' to 'Extreme Annoyance', as shown in Table 5. As previously mentioned, the rating scale is administered to the subjects in its Italian translation, carried out in accordance with the methodology set out in the previous paragraphs. The Italian version used is shown in Table 6.

Verbal Descriptors	Little Annoyance	Moderate Annoyance	High Annoyance	Very High Annoyance	Extreme Annoyance
Category	1-2	3-4	5-6	7-8	9-10

Table 5 1-10 Annoyance Rating Scale

Verbal Descriptors	Poco Fastidioso	Leggermente Fastidioso	Moderatamente Fastidioso	Molto Fastidioso	Estremamente Fastidioso
Category	1-2	3-4	5-6	7-8	9-10

Table 6 1-10 Annoyance Rating Scale, Italian translation.

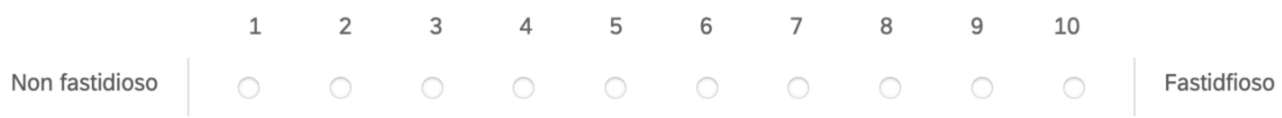


Figure 26 a 1-10 Annoyance Rating Scale used in on the tractor test.

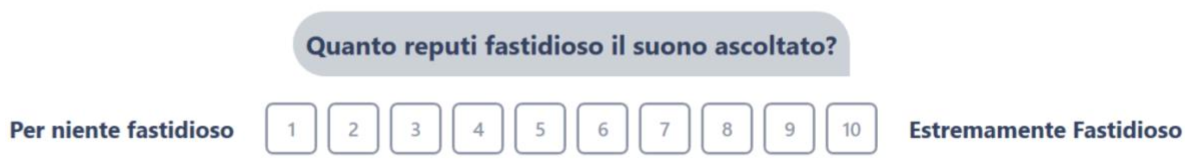


Figure 26 b 1-10 Annoyance Rating Scale used in on the Binaural test.

Figure 26 c 1-10 Annoyance Rating Scale used in on the Immersive tests.

2.3.1.2 Semantic Differential Method

To have a clearer and more complete idea regarding the perception of the various psycho-acoustic parameters by the subjects, the Semantic Differential Method is used. In this way, the tested subjects will have at their disposal, for each psycho-acoustic parameter that they want to evaluate, a pair of bipolar adjectives to which numerical values correspond. The adjectives represent the spectrum of the parameter and were translated with the methodology set out in the previous paragraphs, the results of which are reported in Table 3. For example, the Roughness parameter was translated into two adjectives to be placed at the extremes of the scale, i.e. Not Rough and Rough. In this way, the value -3 represents the perception of a sound with very low roughness and the value 3 indicates a perceived very high roughness.

The parameters chosen for this methodology were chosen from Leite's article, from the results of the bibliographic research and from the specific requests of the Denso company. In Table 3 it is possible to see the translation of the chosen parameters from English to Italian.

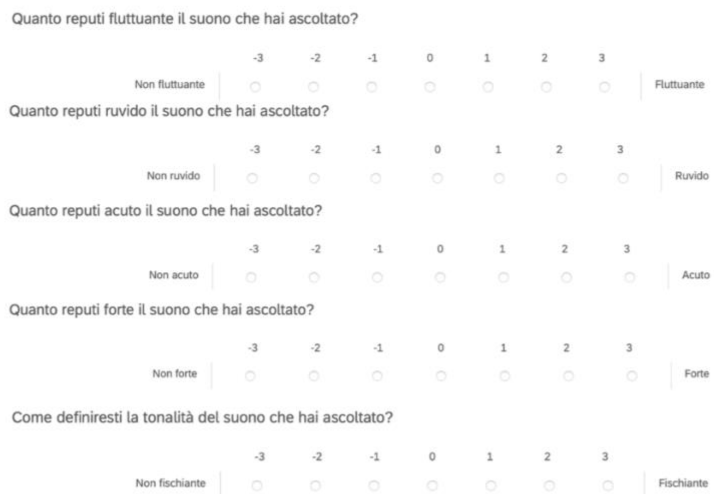


Figure 26d Semantic Differential Method scale used in the tractor test

Figure 26e Semantic Differential Method scale used in the binaural test

Valuta le seguenti caratteristiche sonore dell'audio che hai ascoltato:

Non Forte	-3	-2	-1	0	1	2	3	Forte
Non Ruvido	-3	-2	-1	0	1	2	3	Ruvido
Non Acuto	-3	-2	-1	0	1	2	3	Acuto
Non Fluttuante	-3	-2	-1	0	1	2	3	Fluttuante
Non FIschiante	-3	-2	-1	0	1	2	3	Fischiante

Figure 26f Semantic Differential Method scale used in the Immersive tests

2.3.2 Sound Playback Systems

2.3.2.1 On tractor test

The test on the tractor was supported by 15 subjects of different ages and genders, belonging to the team from the Polytechnic of Turin, who went to the company for the measurements, and to the Denso company team. The subjects took turns getting on the tractor, in which the doors and windows were closed. The HVAC was controlled from the outside in the various powers tested. The subjects had access to a tablet in which to complete the questionnaire consisting of the Rating Scale Method and the Semantic Differential Method.
In this case, there is no sound playback system, but users were subjected to listening to the direct source. The results collected will then be compared with those of other laboratory tests in order to visualize and understand how the playback system influences the perception and responses of the tested subjects.
The subjects subjected to this type of test, being part of different and varied groups, were subjected to a subjective questionnaire, which made it possible to verify the presence of subjects who were experts in acoustics and musicians, in such a way as to highlight how personal background influences on responses and perceptions of sounds. These results will be analyzed in depth in the Results section.

2.3.2.2 Immersive test

The immersive test take place in Audio Space Lab, a laboratory situated in Politecnico of Turin that has 16 loudspeakers that allows to reproduce with high fidelity the sound recorded with microphone arrays. The number of loudspeakers makes the laboratory a third order ambisonics space.

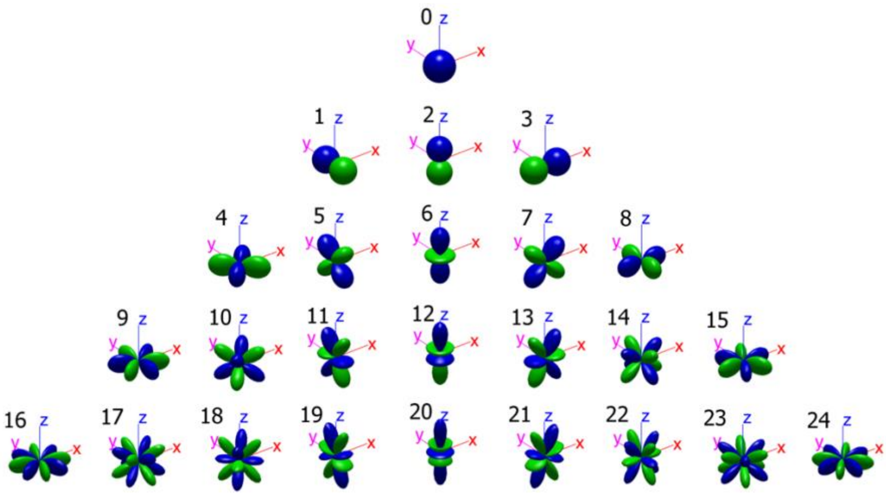


Figure 27 orders of ambisonics

Ambisonic audio is a full-sphere surround sound technology that captures and reproduces sound from all directions around a single point. Unlike traditional stereo or even 5.1/7.1 surround sound, which are based on channel-based audio (where each speaker has a specific sound assigned to it), ambisonic audio is a scene-based approach. This means that the sound is described in terms of the direction it comes from, allowing for a more immersive and natural listening experience, especially when paired with virtual reality (VR) or 360-degree video.

This technology uses spherical harmonics to encode the sound field. The sound is captured using an ambisonic microphone, typically consisting of multiple capsules arranged in a specific geometry. The captured sound is then encoded into ambisonic “channels” or “orders” that describe the sound field around the listener.

The “order” of ambisonics refers to the level of detail in the spatial representation of the sound. Higher orders provide more precise localization and a more accurate representation of the sound field, but they also require more data and processing power. There are three common orders of ambisonics audio.

First-Order Ambisonics (FOA)

- **Number of channels**: 4 channels (W, X, Y, Z)
- **W**: Represents the omnidirectional sound (like a mono signal).
- **X, Y, Z**: Represent the directional components along the front-back, left-right, and up-down axes, respectively.
- **Coverage**: Provides a basic spherical representation of the sound field.
- **Usage**: First-order ambisonics is the most common and widely supported format. It gives a decent sense of directionality but with limited precision, especially for sounds coming from above or below.

Second-Order Ambisonics (SOA)

- **Number of channels**: 9 channels
- Includes all components of first order (W, X, Y, Z) plus 5 additional channels that provide more detailed information about the sound field’s curvature.
- **Coverage**: Offers better spatial resolution than first order, improving the accuracy of sound localization.
- **Usage**: Second-order ambisonics are more precise in representing the sound field, especially in differentiating sounds coming from closely spaced directions. However, it requires more processing power and is less commonly used than first order.

Third-Order Ambisonics (TOA)

- **Number of channels**: 16 channels
- Includes all components of first and second order plus 7 additional channels that further refine the spatial detail.
- **Coverage**: Provides high spatial resolution, allowing for very accurate localization and a more immersive experience.
- **Usage**: Third-order ambisonics is used in high-end applications where precise spatial audio is crucial, such as in professional VR environments, high-quality 3D audio production, and advanced research applications. It requires significantly more computational power and storage.

Higher order ambisonics (beyond third order) exist, further increasing spatial accuracy, but they become increasingly complex and require more resources to process. This allows a good quality reproduction of the low components of the tractor HVAC noise.

To the fifteen subjects is also given a Virtual Reality headset, the Meta Quest 2, in order to make the experience more immersive. The VR headset allows the reproduction of the 360° video recorded in the tractor with the 360° camera Insta 360 X One, shown in **Figure 29** with the Meta Quest 2 headset. The video is monoscopic meaning flat images are projected onto a sphere at 25 frames per second, shown in **Figure 28**.

Figure 28 monoscopic view captured with Insta360 One X2 camera.

Figure 29 Insta360 One X2 and Meta Quest Headset used

In order to have a correct reproduction of the recorders obtained with the Zylia ZM-1 microphone that has 19 channels of acquisition, a convolution was necessary, to take the number of channels from 19 to 16, making them ambisonics. This passage is done with the software Bidule.
The volume of each recording from the Zylia ZM-1 microphone array was manually fine-tuned to align with the LZeq of the NTi omnidirectional microphone recordings taken inside the tractor cabin at the operator's ear level. This calibration process involved placing the NTi microphone at the listener's ear level within the audio space lab, playing back each sample corresponding to the operating mode listed in Table 5, and adjusting the playback volume according to the respective LZeq value.
The script of the immersive test allows for a coordinated test environment involving three programs: Bidule for audio control, Unreal Engine for playing audio stimuli, and MATLAB for facilitating communication between the two through Open Sound Control (OSC). It establishes OSC connections, sets file paths and selects specific sound files. The script manages subject information, manages training stimuli, and prompts users to rate audio stimuli on a scale of 1 to 10. The script collects Rating Scale (RS) and Semantic Differential Method (SDM) ratings for all modes of operation and organizes all collected data into Excel files. It also provides an automated structure for running experiments with audio stimuli, managing subject data and exporting results.

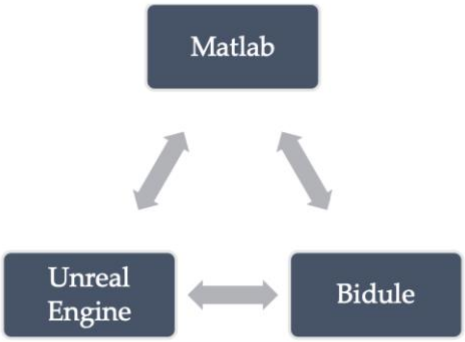


Figure 30 communication between programs for the immersive test.

This test is conducted similarly to Favaretto's research [39] to compare the results and prediction models.

2.3.2.3 Immersive Listening Test with Headphones

Another fifteen subjects were selected to participate in the immersive test, which took place at the Polytechnic of Turin. The test aims to represent the tractor environment and the sound of the HVAC as faithfully as possible. To do this, a virtual reality viewer, the Meta Quest VR Headset, and a pair of high-quality headphones, the Sennheiser 650 HD, were needed. The viewer allows you to play a 360° video and the recorded sound of the HVAC is reproduced in a spatialized manner in the headphones. This is possible thanks to the spatialization of the audio which took place after the recording, which allowed us to obtain a third order ambisonic recording, referencing to **Figure 31**.

In this research, a third order of ambisonics was used, therefore 16 virtual channels, as reproduced in headphones and generated with the FB360 plugin connected to the Reaper software.

The recordings used for this type of test are those obtained with the 19-channel Zylia ZM-1 microphone. The recordings were, as a first step, calibrated, in such a way as to have a volume level and output equal to that recorded by the HSU III.3 artificial head, taken as a reference as the calibration is known. To do this, the headphones used for listening, the Sennheiser 650HD, were placed on the artificial head. From the computer connected to the device, using the Artemis Suite software, the sound recorded by the Zylia microphone was reproduced and a recording of the playback was acquired. The LAeq values were extracted from it and compared with those acquired with the artificial head. Subsequently, the gain of the Zyila ZM-1 microphone track was modified, and the process was repeated until an LAeq level equal to that of HSU III.3 was obtained. This calibration process was carried out in the anechoic chamber present at the Polytechnic of Turin.


Fig. 31 playback volume adjustment in anechoic chamber

Once the correct values were obtained, we proceeded to perform the spatialization of the channels acquired by the Zylia. To do this, it was necessary to have installed the FB360 plugin for Reaper, a Meta-owned extension that allows you to manipulate the audio and make it the ambisonic order you need. The software allows, by displaying a sphere around the listener, to position the individual sources in space. In **Figure 32** there is an example of what the interface used to spatialize the audio looks like.

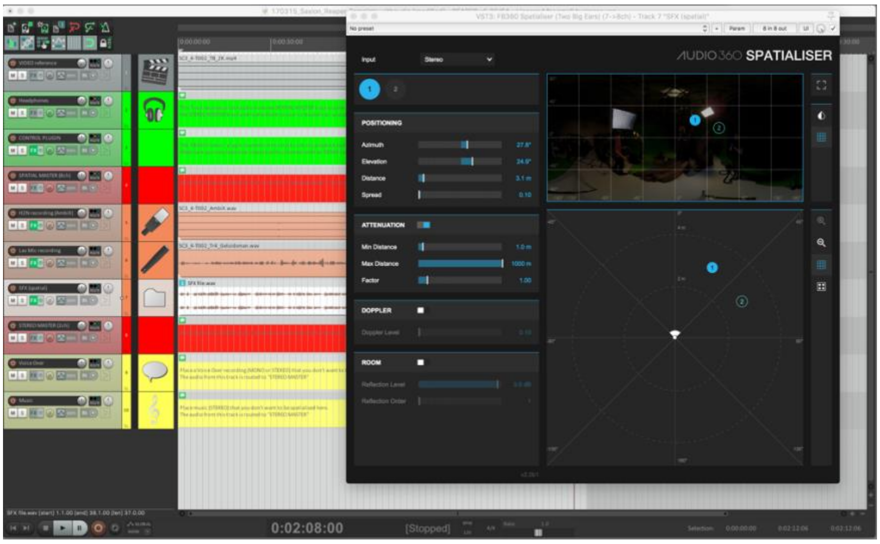


Figure 32 FB 360 Spatializer Workstation's Interface

The 360° video was recorded through the Insta 360 One of the tractor. The recording format is monoscopic, meaning flat images are projected onto a sphere at 25 frames per second.
Finally, the spatialized audio was mounted on the 360° video, in such a way as to obtain immersive multimedia content that could be used with the virtual reality viewer.
To reproduce the content, mirroring on the PC monitor was used, to manage the reproduction and display of the scales. The responses were recorded in an Excel file and subsequently analyzed.

2.3.2.4 Binaural test

The binaural test involved another 15 different subjects, who were subjected to 10-second samples extracted from the recordings carried out with the HSU III.3 artificial head. The test was carried out in the anechoic chamber of the Polytechnic of Turin and Sennheiser HD 650 headphones, a laptop and the HeadAcoustics home software, Artemis Suite, were used. In particular, the SQala module was used to create the questionnaire and allow the reproduction of the audio.
The subjects, therefore, wore headphones and answered the questions independently.
No subsequent correction or calibration was necessary, as the correct parameters were set during registration. The ID equalization filter, or Independent Direction, was chosen to record as the characteristics of the vehicle deemed it the most suitable. Furthermore, this type of artificial head does not have an auricle, but everything is managed via software.

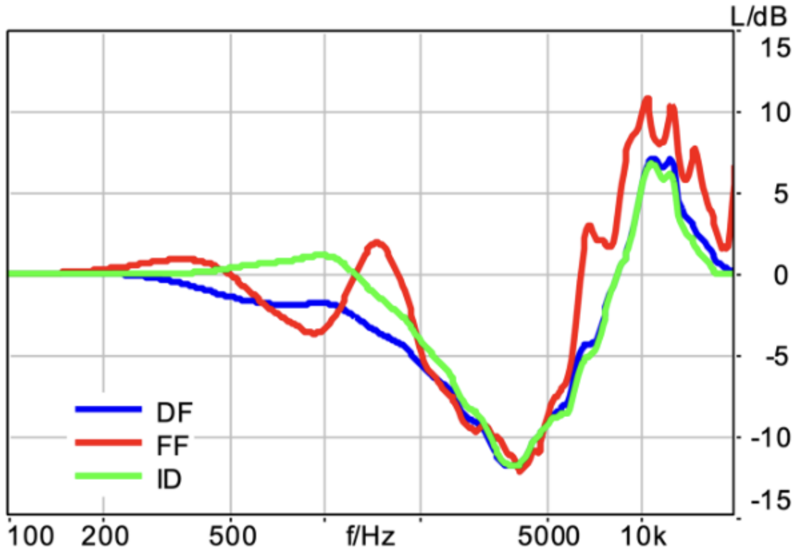


Figure 33 independent-direction, diffuse field and free field equalization curves

For these reasons, the recordings obtained with the artificial head were taken as reference for the calibration of the other microphones.

3. Results

In this chapter, the psycho-acoustic parameters extracted from the recordings acquired on the tractor, the results of the subjective tests and the calculated predictive models will be analyzed in depth.
Ten seconds of audio were extracted from the recordings without interference or external noise and on these the calculations of the psycho-acoustic parameters and the subjective tests with Rating Scale Method and Semantic Differential Method were performed, as explained in the previous paragraphs.
To calculate the predictive models, the multiple linear regression method was chosen. The averages calculated on the responses to the subjective tests represent the dependent values. The averages of the psychoacoustic parameters calculated from the recordings appear to be the independent values.
The primary parameters included are the Equivalent Sound Pressure Level (LZeq), the A-weighted Equivalent Sound Pressure Level (LAeq), Loudness, Sharpness, Fluctuation Strength, Tonality and Roughness. Supplementary material also includes Speech Interface Level and Articulation Index.
The analysis of the parameters and subjective tests and the linear regression were carried out with the same methodologies used by Marco Favaretto in his research [39], in such a way as to have comparable and comparable results. The comparison between the results obtained with the two methodologies will be explored in depth in the following paragraphs.

3.1 Psycho acoustic Analysis

The recordings analyzed for the evaluation of the objective parameters are those acquired with the HSU III.3 artificial head and with the NTi omnidirectional microphone. The recordings were cut to a length of 10 seconds, taking care to choose intervals such that there were no noise or external interference. The same portions of audio are those proposed to subjects subjected to subjective tests.
The objective parameters were extracted with the analysis module of the Artemis Suite software. The calibration relating to the artificial head was automatically loaded by the software and that relating to the NTi microphone was manually loaded into the software.

The parameters were extracted with the settings listed below:

Acquisition Parameters: insert > new track > mark editor > cut track to 10 seconds length > insert: Loudness ISO 532A, Sharpness - diffuse field, Roughness, Fluctuation Strength, LAeq – A – weighted equivalent continuous sound level, Level – Unfiltered, Tonality, Speech interference level, Articulation index, FFT vs Time, Frequency Spectrum > Mark Analyzer.

These settings are the same ones used in Marco Favaretto's research [39], which allows for an accurate and consistent comparison of results.
In order to have a better picture of the characteristics of the noise, parameters strictly related to the spectrum of the sound are extracted and analyzed, like Noise Criterion, Room Noise Criterion, Noise Rating and Preferred Noise Rating.

3.1.1 Binaural and Omnidirectional Microphones

Four ten-second audio extracts acquired with the HSU III.3 artificial head were chosen from all the recorded samples, one for each HVAC speed considered, and the same number for the NTi omnidirectional microphone. The extracts from the artificial head are the same ones that were administered to the subjects involved in the laboratory tests.
The subjective ratings obtained will be utilized as the dependent variable, while the psychoacoustic results will serve as the independent variables for the prediction model.
Below is the comparative analysis of the results obtained from the two microphones.

3.1.1.1 Equivalent Sound Pressure Level

The LAeq parameter was measured in dBA
Lzeq levels increase with increasing speed, for the head with a variation of approximately 15dBSPL on the left and 18dBSPL on the right. As regards the head, there is no big difference between left and right, with a slight imbalance of around 2 dBSPL in favor of the right side.
The NTI records lower levels but the trend is consistent with the data recorded from the head.
A difference of approximately 13dBSPL was recorded between Background and Vel12.

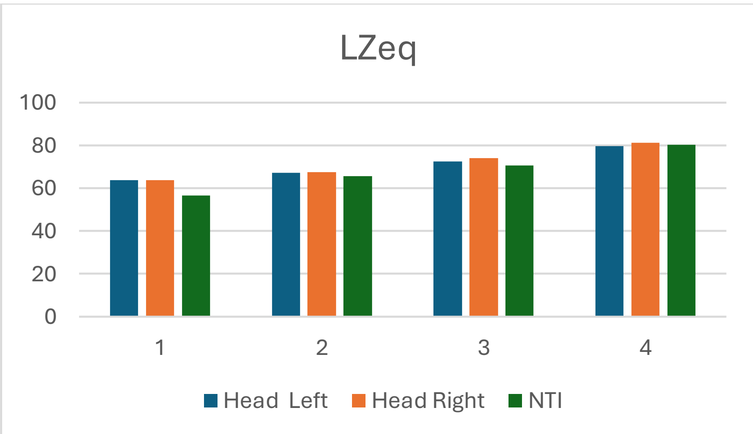


Figure 33 (y-axis) noise equivalent sound pressure level expressed in dB; (x-axis) type of measurement from. JND=1dB

3.1.1.2 A-weighted Equivalent Sound Pressure Level

The LAeq parameter was measured in dBA
Laeq levels increase with increasing speed, with a difference between background and vel12 for the head of approximately 40dBSPL(A) on the left and 41dBSPL(A) on the right.

For the head, the difference between left and right in this case is below the jdn threshold except for velocity4, unbalanced by just over 1 dBSPL(A) to the left
The NTI also in this case has an overall level recorded lower than the head, but with a trend consistent with it. An overall difference of approximately 24 dBSPL(A) was recorded between Background and speed 12.
We note how the difference between speed 8 and speed 12 is smaller than that between the other speeds examined, which did not occur with LZeq.

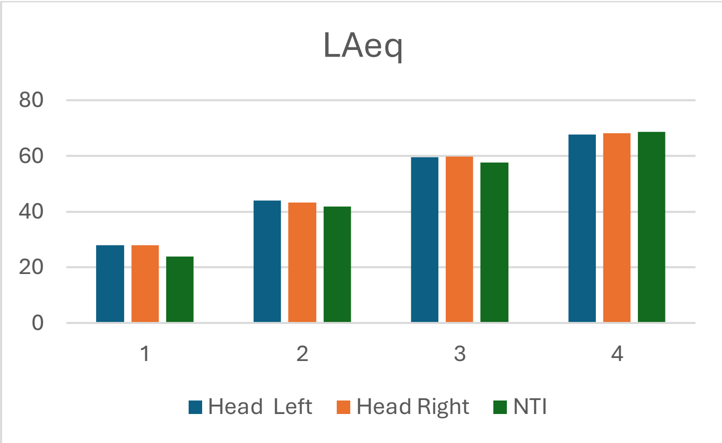


Figure 34 (y-axis) noise A-weighted sound pressure level expressed in dBA; (x-axis) type of measurement JND=1dB

3.1.1.3. Loudness

Loudness was measured in sones.
Loudness levels grow almost exponentially as speed increases, with a difference between background and head speed of 23 sones on the left and 24 sones on the right.
The levels recorded with the head do not record differences between right and left greater than the jnd, therefore not perceptible to the human ear.
The NTI has a much lower overall recorded level than the head, with recorded values equal to approximately half of the head. However, the trend is consistent with the other microphone.
The difference between background and speed 12 recorded by the NTI is approximately 12sone.
The substantial differences for both microphones were recorded between speed 4 - speed 8 and between speed 8 - speed 12, while between background and speed 4 the difference does not exceed jnd.
The same thing happens for the differences between microphones, in the background and at speed 4 the recorded values do not present significant variations.

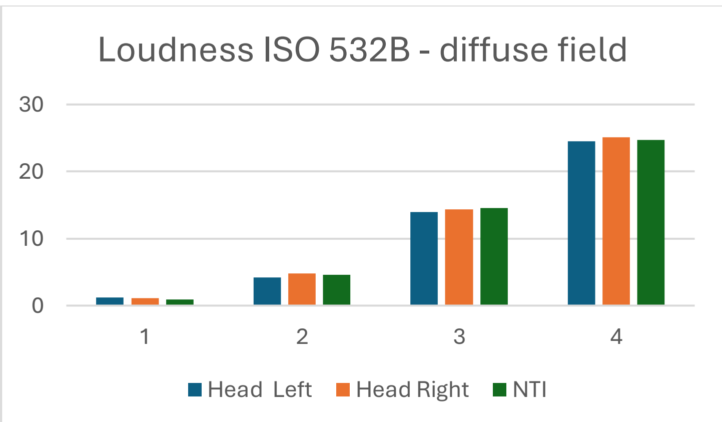


Figure 35 (y-axis) mean Loudness level expressed in sone; (x-axis) type of measurement from JND=0.8 sone

3.1.1.4. Sharpness

Sharpness was measured in acum.
Sharpness levels increase as the HVAC speed increases, with a difference between background and vel12 for the head of 1.2 acum on the left and 1.3 acum on the right.
The levels recorded with the head show perceptible differences between right and left regarding the background and speed 4, while for 8 and 12 it is below the jnd.
The NTI records overall lower values than the head, with a difference that increases as the speed increases. In fact, we go from a difference of 0.1 to a difference of 0.4.
Overall, the trend recorded by the NTI is consistent with that of the head, with an overall difference between speed 12 and background of just over 1 acum.
There is a substantial difference in the sharpness levels recorded between speed 4 and speed 8, with approximately a doubling of the acum value, while between speed 4 - background and speed 8 - speed 12 the difference is more limited.

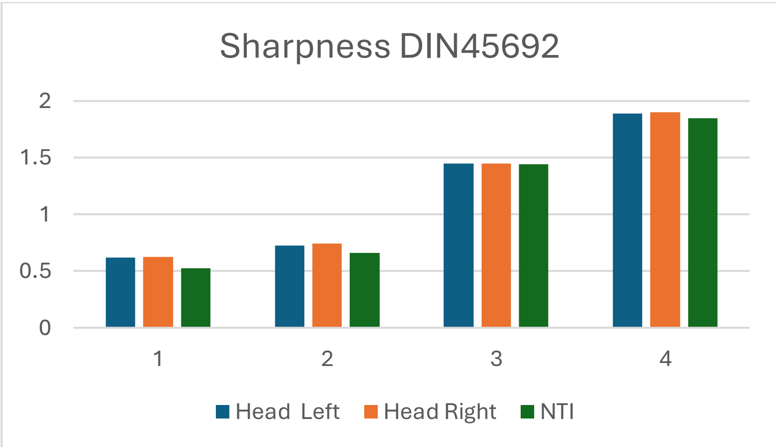


Figure 36 (y-axis) mean Sharpness level expressed in acum; (x-axis) type of measurement from JND= 0,04acum

3.1.1.5. Roughness

Roughness was measured in asper.
The roughness does not increase as the air outlet speed increases but has a peak at speed 4 and then decays as the speed increases.
The head recorded differences between the right and left values smaller than the jnd, with an overall difference between speed 12 and background of approximately 0.1 asper on the right and left.
The NTI recorded generally lower values than the head and the trend is not consistent with that of the HEAD 3.III, with the roughness value increasing with the speed of the HVAC.
a difference between background and vel12 of approximately 0.08 asper was recorded with the NTI.
There is therefore a discrepancy between the trend of roughness recorded by the head and by the NTI.

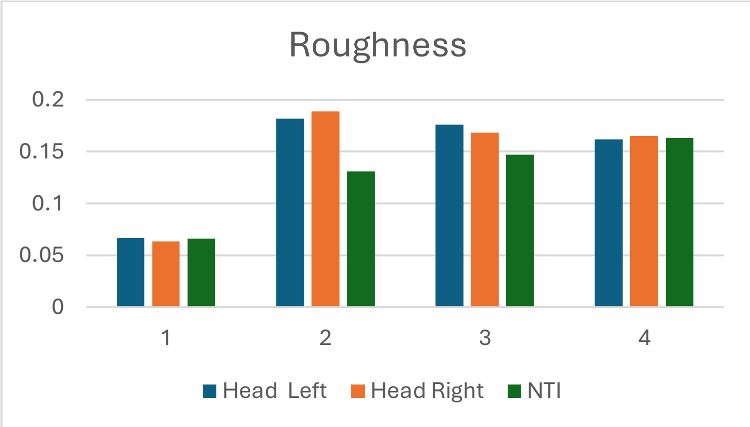


Figure 37 (y-axis) mean Roughness level expressed in sone; (x-axis) type of measurement from JND=0.04 asper

3.1.1.6 Fluctuation Strength

Fluctuation strength was measured in vacil.
The fluctuation strength recorded with the head increases as the HVAC speed increases.
The difference between the right and left channels is very low, it does not approach the jnd threshold, so the two channels are perceptually identical.
The difference recorded between the background and the 12 speed is approximately a little less than 0.01 on the left and right, therefore with a smaller overall difference compared to the jnd.
The NTI recorded lower values than the head, but with a trend consistent with it. The difference between the two microphones is less than the jnd for each speed.
The difference between the background and the 12-speed recorded with the NTI is approximately 0.005, well below the jnd. Note how there is a doubling between background - speed 4 and between speed 4 - speed 8, while between speed 8 and speed 12 the difference is minimal.

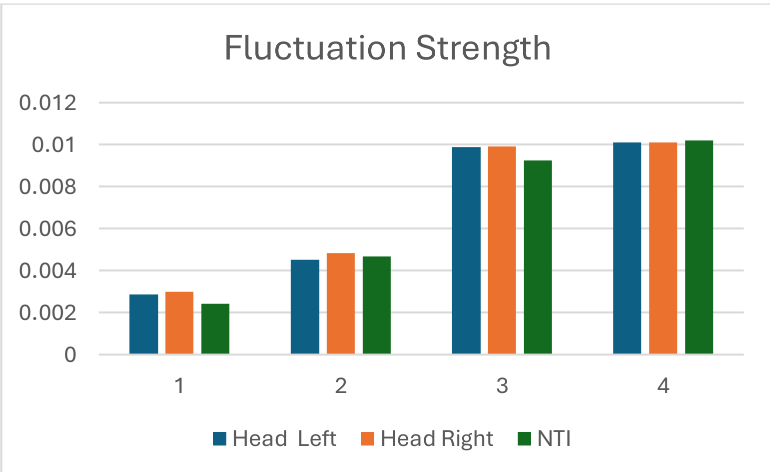


Figure 38 (y-axis) mean Sharpness level expressed in vacil; (x-axis) type of measurement from JND=0.012 vacil

3.1.1.7 Tonality

Tonality was measured in tuHMS.
The tonality recorded by the head does not have a trend consistent with the speed of the HVAC. There is a peak with speed 4, a decrease with speed 8 and a new peak with speed 12.

The difference between the right and left head channels is greater than the jnd and is particularly high in speed 4 with a difference of about 0.1 greater on the left. Even at speed 12 there is a significant disparity of about 0.05 between the left and right channels. The NTI has recorded levels lower overall than the head, with a trend not consistent with that of the head 3.III.In fact, there is an increasing trend as the speed with the NTI increases.
The overall difference between speed 12 and background is approximately 0.2 tuHMS, while for the head there is an overall difference equal to 0.25 both on the right and on the left.
Note how the tonality value recorded with the NTI is approximately 0.02 higher in the background compared to speed 4.

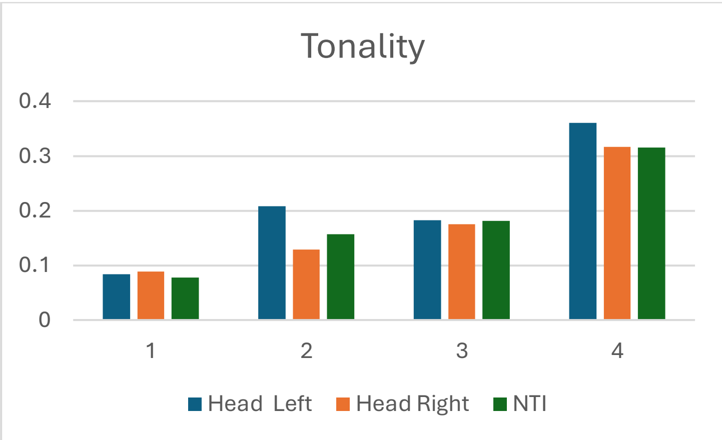


Figure 39 (y-axis) mean Tonality level expressed in tuHMS; (x-axis) type of measurement from JND=0.005 tuHMS

3.1.1.8. Frequency Spectrum

The frequency spectrum is a graph that displays the dB level corresponding to each frequency. This therefore allows us to understand the influence of some specific frequencies.
It can be seen in **Figure 40** and **Figure 41** that the general trend of the artificial head HSU III.3 spectrum is similar for the right channel and the left channel. Furthermore, a similarity in the trend, in the position of the peaks and in the shape of the spectrum can be noted between the different HVAC speeds.
We can see, in the general trend of the curves, that we have much higher dB levels for low frequencies, with a common peak at speeds 4, 8 and 12 at around 100Hz and subsequently one at around 300/400Hz.
A decay begins just above 500Hz and becomes more sudden after the slight peak recorded at 2kHz.
The background and especially the Vel4 have a peak between 20 and 50 Hz at a rather high dBSPL level, equal to approximately 65dBSPL.
The left channel for speed 4 presents a significant peak at high frequencies between 5kHz and 10kHz.
The maximum peak is touched by the blue curve of speed 12 with a value equal to 70dBSPL at 100Hz.
There is a difference between the trend of the spectrum at high frequencies (just before 1kHz) regarding speed 4 compared to 8 and 12. The last two are rather consistent with each other in terms of decay, peaks and shape, while the 4 has a lower trend between 1000 and 5000Hz, then rises with a peak and decays. In the right channel we can see how speed 8 has a peak that exceeds the value of the speed 12 curve between 50 and 100 Hz. This peak is not present as accentuated in the left channel.
Therefore, we can see how the spectrum is rich at very low frequencies (0-50Hz), has its peak at low frequencies (50-500Hz) and decays at high frequencies, with a minimum for the background at around 7kHz.
On a general level the two spectrums are quite similar, with slight variations regarding left and right.
Finally, we can say that the high frequency component is different and more irregular for speed 4 compared to 12 and 8, as well as for the very low frequency component.

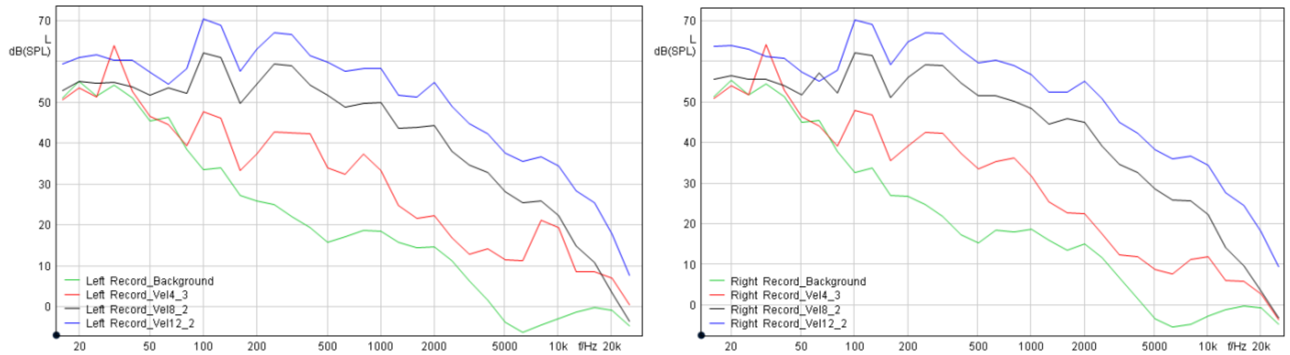


Figure 40 – 41 Spectrum Analysis of the artificial head HSU III.3, depicting the left (40) and right (41) channels. Dashed lines indicate HVAC speeds with the tractor engine running.

However, as regards the NTi omnidirectional microphone, the spectrum appears to be consistent in terms of trend and peaks with that recorded by the head channels.
We can see how speeds 4, 8 and 12 have a similar trend, especially 12 and 8. Speed 4 has a peak between 20 and 50 Hz that exceeds 50 dBSPL, which is not present for speeds 8 and 12. Furthermore, speed 4 presents a trend at high frequencies with a steeper decay and with a peak between 5 and 10 kHz does not present in the other spectra. The background is irregular, with levels above speed 8 at very low frequencies and a significant decay between 200 and 1000 Hz. Speed 8 has an isolated peak between 50 and 100 Hz, which exceeds the value of speed 12 in terms of dBSPL.
The maximum peak is reached by speed 12 between 100 and 200Hz with a value of approximately 60dBSPL.
We can see, in the general trend of the curves, that we have much higher dB levels for low frequencies, with a common peak at speeds 4, 8 and 12 at around 100Hz and subsequently one at around 300/400Hz.
Therefore, this spectrum also presents richness between 0 and 100Hz, with significant peaks between 100 and 500 Hz, and then decays after 500 Hz, more suddenly after 2kHz.
We can then say that the high frequency component is different and more irregular for speed 4 compared to 12 and 8, as well as for the very low frequency component.
Finally, we can say that the NTI spectrum is consistent with that of the Head but has a lower dBSPL level.

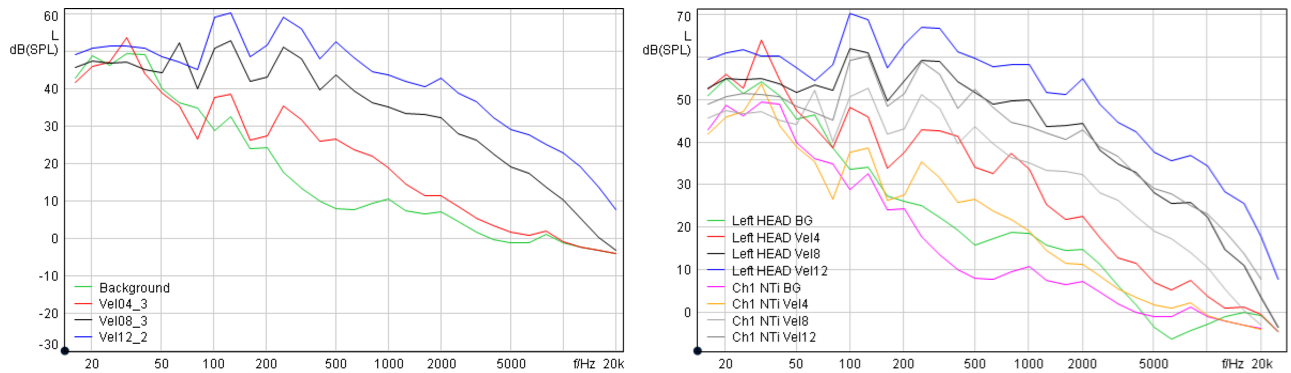


Figure 42 – 43 Spectrum Analysis of the omnidirectional microphone NTI (42) and overlapping spectrum analysis of the NTi and HSU III.3's Left Channel(43) channels. Dashed lines indicate HVAC speeds.

3.1.1.9. FFT vs Time

The analysis of FFT vs Time allows you to see how the frequency distribution changes over time.
As regards the background recording, the left and right channels of the artificial head are quite similar, with a strong low frequency component, up to around 50 Hz, which then attenuates between 50 and 100 Hz and then becomes quite low between 100 and 300 Hz and then almost zero.
We have some different peaks at low frequencies over time between the left and right channels.
At 100 Hz the left channel has a much more present and constant component than the right. On the left there is an isolated component at 200 Hz at 110 t/s.

There is also blurring at a dB level corresponding to approximately 20-25 between 700 and 2000Hz, both on the right and on the left, but distributed differently over time.
The NTI spectrum is slightly different. It has a strong low frequency component, up to 50Hz, drops before 100Hz, decays around 200Hz and then becomes almost imperceptible.
There is a time-dilating component between 100 and 200Hz that lasts approximately 100 to 150t/s at a dB level between 40 and 50. There are other isolated components of this type, also present in the head but not as prominent and prolonged.

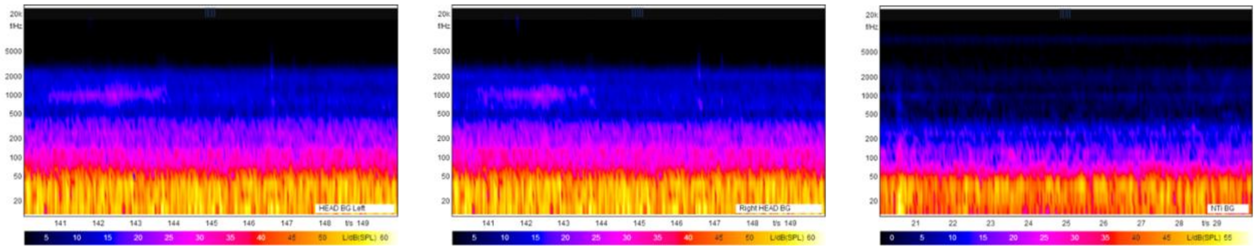


Figure 44 – 45 – 46 FFT vs Time of the artificial head HSU III.3, depicting the left (44) and right (45) channels, and NTi microphone (46). BG is the current operating mode.

Analyzing the operating speed 4 of the HVAC, we can see how the right and left spectrum of the head are rather aligned, with a strong low frequency component (0-50Hz) almost constant over time. There are peaks at around 35 and 45 t/s in both spectra. Between 50 and 100 Hz the level drops, with a stronger level at 50 and 100 Hz and quieter in between. We have some burrs between 20 and 25 t/s and between 25 and 35t/s.
Going up with the frequencies we have a decrease in dB between 100 and 500Hz, with an average level of around 30/35dB, but with some more predominant frequencies corresponding to the pink horizontal lines, at around 200Hz, just before 500Hz and halfway between 200 and 500Hz. We can see how on the left the 500Hz component is very present and defined throughout the recording, while on the right the component is not too perceptible. Above 1000Hz nothing significant is recorded.
The NTI has a much more pronounced low frequency component (0-50Hz) throughout the recording. Even at 100Hz much higher levels were recorded. Between 50 and 100Hz the level drops and also between 100 and 200Hz, with more prominent frequencies. Between 200 and 500Hz we have the same behavior as the head, with a stronger presence at around 300Hz and a defined and specific component at around 400/450Hz. Then we decrease the volume up to 1000Hz, above which we have nothing significant.

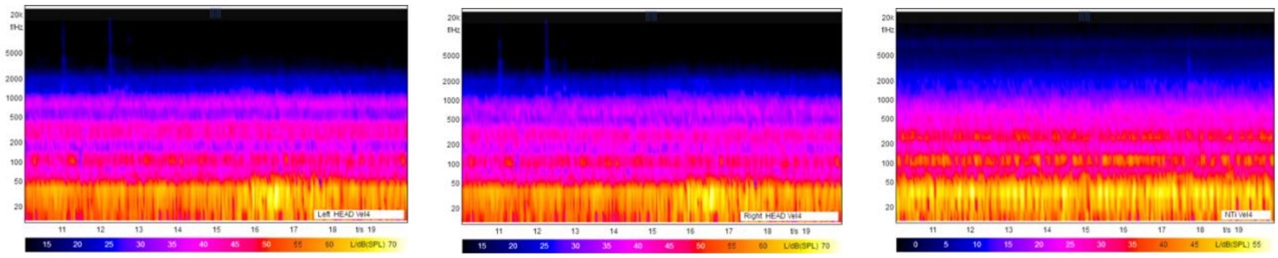


Figure 47 – 48– 49 FFT vs Time of the artificial head HSU III.3, depicting the left (44) and right (45) channels, and NTi microphone (46). Vel4 is the current operating mode.

As regards speed 8 of the HVAC, the low frequency spectral component is less constant than speed 4 and the background, we have a more fragmented trend over time between 50 and 0 Hz. Between 50 and 100 Hz we have a decrease and then you have a strong spectral component at 100 Hz.
Between 100 and 200Hz the spectral presence decreases inconsistently and then has a strong component again just above 200Hz. At 500 Hz there is a significant component and between 500 and 100 there are isolated components. Like the one at 1000Hz.
On the right, the component between 500 and 1000Hz is more defined than on the left, but that at 1000Hz is decidedly less so. Above 1000Hz the general level drops and then has no significant component above 2000Hz.
With the NTI we have a similar behavior, the low frequency component is more prominent than with the head.

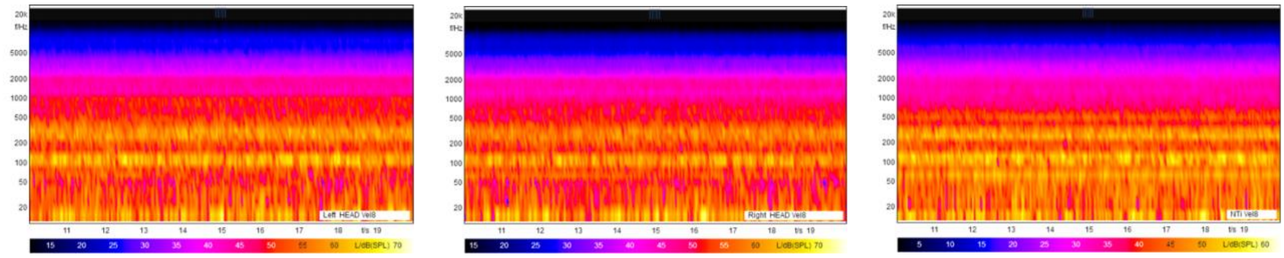


Figure 50 – 51– 52 FFT vs Time of the artificial head HSU III.3, depicting the left (44) and right (45) channels, and NTi microphone (46). Vel8 is the current operating mode.

Analyzing the spectrum of speed 12, we can see how there are no large variations between left and right, the low frequency component is similar to speed 8, therefore less constant and more jagged over time. We have a sag between 50 and 100Hz and a strong component around 100Hz.
Between 200 Hz and 500Hz we have a rather strong component, it seems around 300Hz. At 500Hz there is a lowering and we can notice the presence of more defined frequencies on the left as at 1000Hz and a little below. On the right we have a very defined component just above 500Hz and there are no defined components above, but there is a significant level between 500 and 1000Hz.
The NTI has a similar pattern, with a more present and constant low-frequency component for the duration of the recording. There is a stronger 100Hz component than with the head, and then lowers to 200Hz.
We have another very present component just above 200Hz and then it lowers up to 500Hz, where we have another component defined around 50-55dB. From 500 to 2000Hz we have the presence of defined components for the entire recording, as can be seen from the fuchsia lines representing 40-45dB.

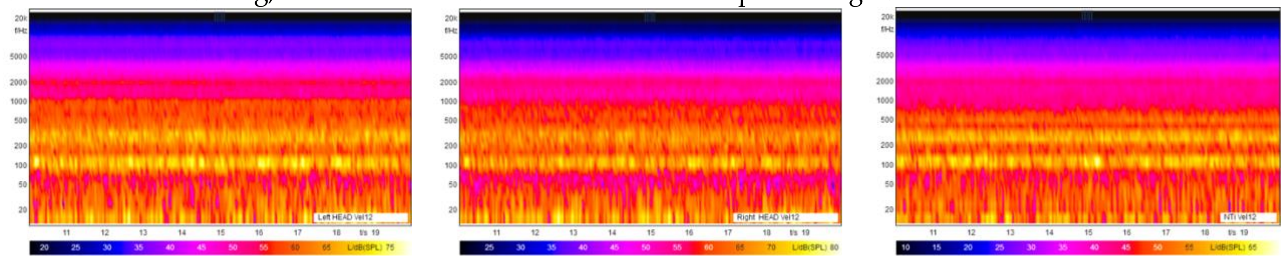


Figure 53 – 54– 55 FFT vs Time of the artificial head HSU III.3, depicting the left (44) and right (45) channels, and NTi microphone (46). Vel8 is the current operating mode.

To conclude, as we can see from the graphs, it is clear that the analyzed signal is stationary, as the dB level remains almost constant over time. We can say that the low frequency component becomes less predominant by increasing the speed of the HVAC, therefore with speed 8 and 12.
On the other hand, the highest frequency component becomes more relevant and increases, reaching 2000Hz of relevant components. Finally, we can say that the most important and common component at all three speeds is found at 100Hz and becomes more important as speed increases, and even just above 200Hz and 500Hz.
Therefore, also in agreement with the filtered frequency spectrum, we can say that there is a strong predominant component at low frequencies and is poorer at high frequencies. It is also noted that the peaks of the two spectra are absolutely in agreement.

3.1.1.10. Supplementary material

The articulation index is a parameter that allows the intelligibility of speech to be quantified through a comparative analysis of speech and background levels. This parameter is measured as a percentage and the higher the value, the better speech clarity we have.
The value is very high for the background and for speed 4 of both microphones, which shows that there is not a noise value high enough to prevent speech.
The value drops drastically for the head with speed 8 and is halved with speed 12

The NTI records much higher values for both speed 12 and speed 8.
There is not much discrepancy between the left and right values recorded by the head.

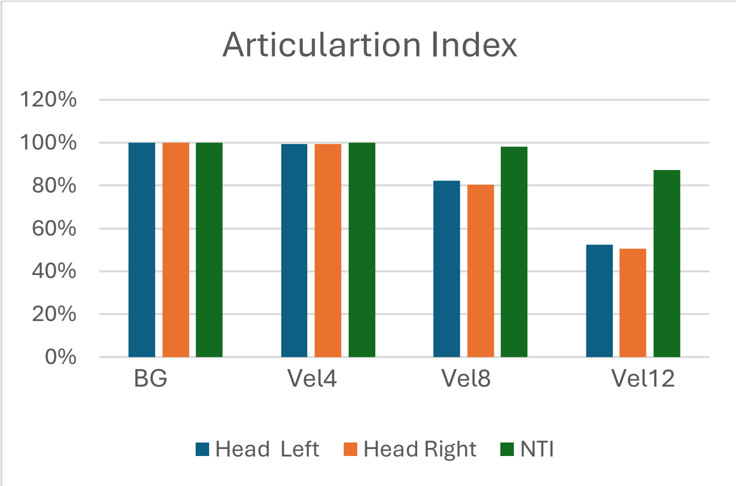


Figure 56 (y-axis) Articulation Index level expressed in percentage; (x-axis) type of measurement from

The Speech Interference Level (SIL) measures the amount of interference caused by background noise on speech communication.
The value of the parameter is very low for the background and for speed 4, in agreement with the values of the articulation index recorded with the artificial head. The value rises significantly for speed 12 and speed 8, in accordance with the articulation index values. . In this case for the head the imbalance is to the right, but as for the Articulation Index there is not a very high discrepancy.
The values recorded by the NTI are much lower for speeds 12 and 8, which agrees with the previous parameter. The microphones are consistent with respect to the two parameters.

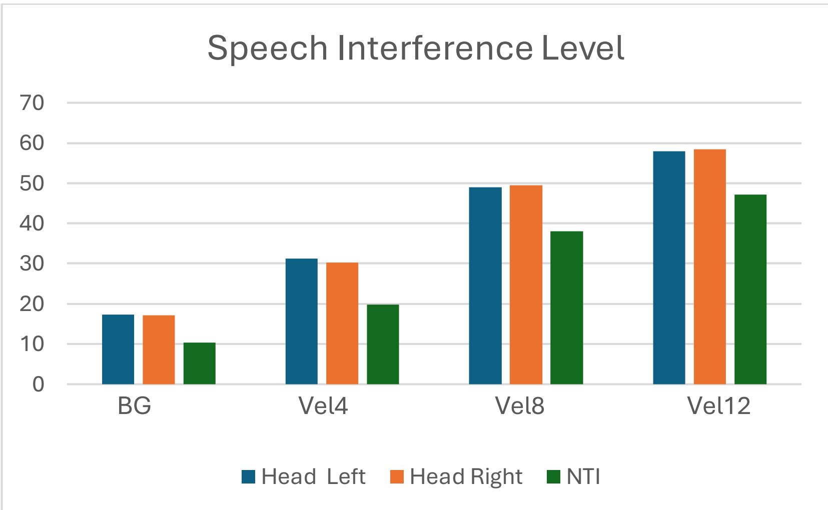


Figure 57 (y-axis) Speech Interface Level expressed in percentage

3.1.1.11 Noise Rating

Noise Rating is a parameter used for classifying environments based on their acoustic suitability.
The trend of the parameter increases with the increase in speed. The difference between the right and left channels, regarding the results obtained with the Head HSU III.3 artificial head, is not significant. The values recorded with the omnidirectional microphone NTi and HSU III.3 show a difference in levels, but the trend remains consistent.

For the background, the recorded values are 18 for the left channel, 20 for the right channel, and 9 with the NTi microphone, all of which fall within the category of very quiet environments. At speed 4, the values recorded with the artificial head fall into the category of moderately quiet environments, while the NTi values are in the category of quiet environments.
The values for speed 8 and speed 12 recorded by the artificial head belong to the category of very noisy environments.
For the NTi microphone, the recordings indicate noisy environments at speed 8 and very noisy environments at speed 12.

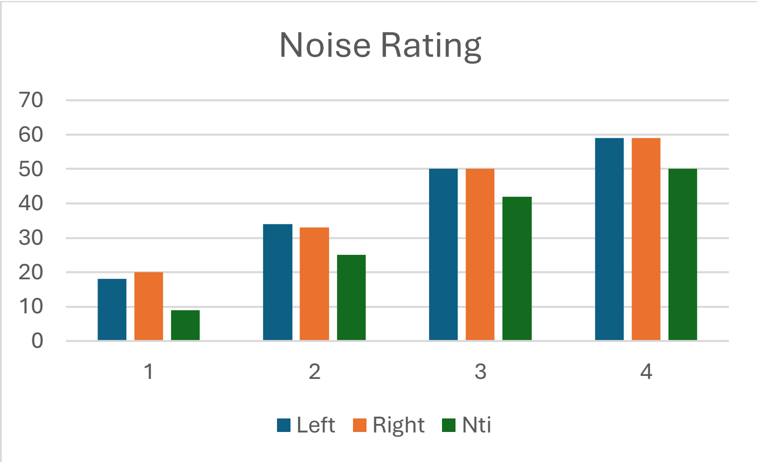


Figure 58 (y-axis) Noise Rating

3.1.1.12 Preferred Noise Criterion

Preferred Noise Criterion is used to evaluate and classify noise in closed spaces, considering human preferences and acoustic comfort.
The trend of the parameter increases with the increase in speed. The difference between the right and left channels, regarding the results obtained with the Head HSU III.3 artificial head, is not significant. The values recorded with the omnidirectional microphone NTi and HSU III.3 show a difference in levels, but the trend remains consistent.
For the background, the recorded values are 19 for the left channel, 20 for the right channel, and 8 with the NTi microphone, all of which fall within the category of very quiet environments. At speed 4, the values recorded with the artificial head fall into the category of normal environments, while the NTi values are still in the very quiet environments category.
The values for speed 8 and speed 12 recorded by the artificial head belong to the category of extremely noisy environments. For the NTi microphone, the recordings indicate noisy environments at speed 8 and extremely noisy environments at speed 12.

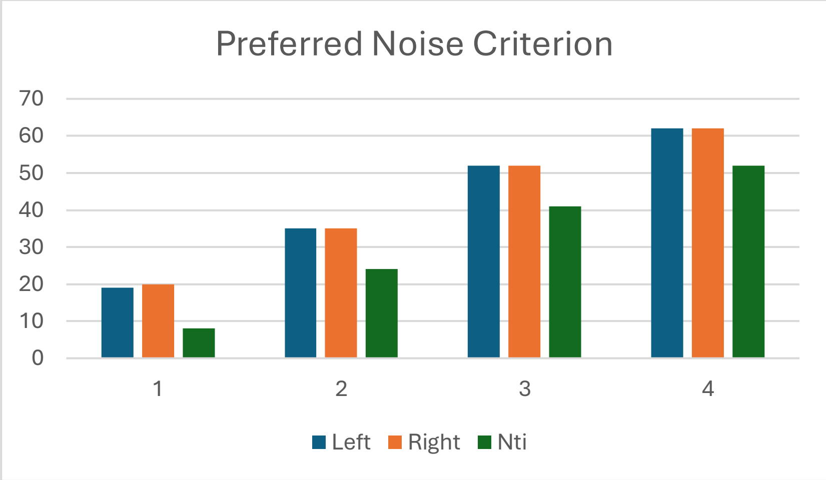


Figure 59 (y-axis) Preferred Noise Criterion

3.1.1.13 Noise Criteria

Noise Criteria is a parameter, based on standardized curves, used for evaluating noise in closed spaces, taking into account the spaces functionality and the acoustic comfort.

The trend of the parameter increases with the increase in speed. The difference between the right and left channels, regarding the results obtained with the Head HSU III.3 artificial head, is not significant. The values recorded with the omnidirectional microphone NTi and HSU III.3 show a difference in levels, but the trend remains consistent.

For the background, the recorded values are 16 for the left channel, 15 for the right channel, and 5 with the NTi microphone, all of which fall within the category of very quiet environments. At speed 4, the values recorded with the artificial head fall into the category of quiet environments, while the NTi values are still in the very quiet environments category.

The value for speed 8 recorded by the artificial head belongs to the category of noisy environments, and the value for speed 12 is labeled as very noisy. For the NTi microphone, the recordings indicate normal environments at speed 8 and very noisy environments at speed 12.

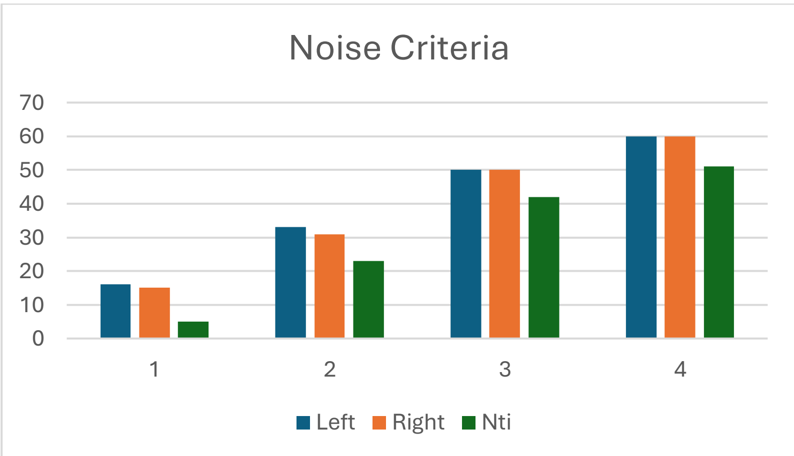


Figure 60 (y-axis) Noise Criteria

3.1.1.14 Room Noise Criterion

Room Noise Criterion is a parameter use for evaluating the noise in a closed space.

The trend of the parameter increases with the increase in speed. The difference between the right and left channels, regarding the results obtained with the Head HSU III.3 artificial head, is not significant. The values recorded with the omnidirectional microphone NTi and HSU III.3 show a difference in levels, but the trend remains consistent.
For the background, the recorded values are 23 for the left channel, 22 for the right, and 11 with the NTi microphone, all of which fall within the category of very quiet environments. At speed 4, the values recorded with the artificial head fall into the category of quiet environments, while the NTi values are still in the very quiet environments category.
At speed 8, the values for the artificial head belong to the noisy environments category, and at speed 12, they fall into the very noisy environments category. For the NTi microphone, the recordings at speed 8 indicate normal environments, while at speed 12, they indicate noisy environments.

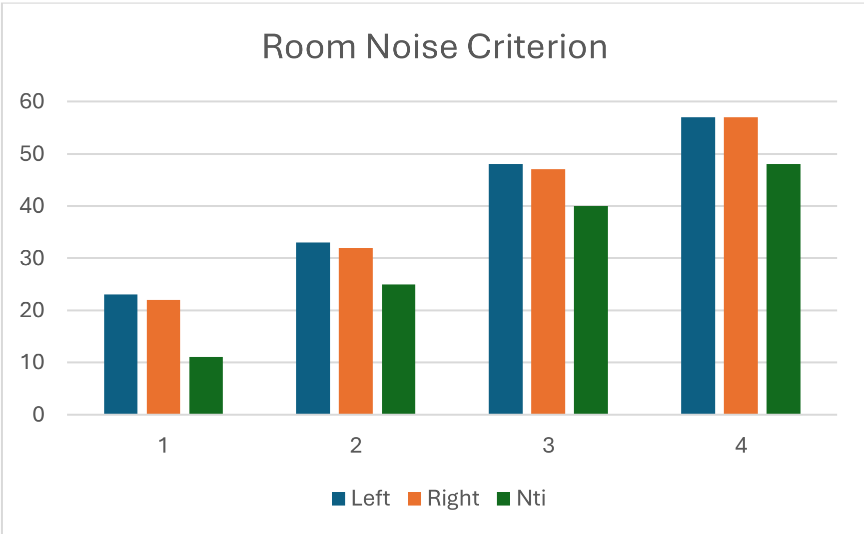


Figure 61 (y-axis) Room Noise Criterion

3.1.1.15. Results Comparison

The results shown in the previous paragraphs appear to be similar and coherent with those analyzed and recorded by Marco Favaretto in his research [39]. The major differences are evident in the analysis of FFT vs Time, which presents a different distribution, but still with similar general characteristics. The same applies to the frequency spectrum.
The parameters based on spectral characteristics show very similar results to one another and the overall trend is coherent with the other parameters taken in analysis.

3.2 Subjective Assessment

The analysis and comparison of subjective responses was conducted for all tests performed during the research. The primary objective is to identify and understand the correlation between subjective responses and objective values calculated with microphones. This lays the foundation for calculating linear regressions, making the subsequent computation more accurate, precise, and reliable. Furthermore, it is also important to assess the differences between responses provided in the various tests to understand which test yields the most reliable responses, particularly in terms of results and alignment between subjective and objective responses. These differences will be further investigated later with the results of the linear regressions.
The analysis was performed on all subjective responses, and subsequently, a subset of responses from subjects with expertise in acoustics and music was extracted, allowing for an evaluation of how prior knowledge influences responses and distributions.

3.2.1. On Tractor Test

The tractor test was conducted at the Denso facility in Poirino, with the participation of 15 subjects who sat on the tractor and listened to and evaluated the noise produced by the HVAC system of the tractor under examination. This test did not require playback system mediation; thus, it was taken as a reference point and benchmark for the accuracy and reliability of the results of the other tests.
The subjects tested were 77.2% men and the age varies from 24 years old to 55 years old. All of the subjects are Italian native speaker and do not have hearing problems.
The temperature changed during the test, from the lower point of 28.5°C to the highest of 37.7°C, as shown in **Table 7a**.
The results obtained from the tractor test were compared with those from both the NTi omnidirectional microphone and the HEAD HSU III.3 artificial head.
Table 7b shows the average values of the responses obtained in the subjective test for the analyzed speeds and parameters.

Subject	Temperature
1	33.75
2	35.9
3	29
4	30
5	33.9
6	37.7
7	32.25
8	33.45
9	35.55
10	31.4
11	29
12	31.5
13	28.5
14	32.8
15	34.4

Table 7a For each subject, the temperature registered inside the tractor cabin.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1.8	-1.5	--1.3	-1.2	-0.6	-1.1
Velocity 4	4.2	-0.7	-0.5	0	-0.5	0.2
Velocity 8	4.9	0.6	-0.2	-0.9	0.1	-0.9
Velocity 12	6	1.5	0.1	-0.3	0.8	-0.1

Table 7b For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

3.2.1.1. Rating Scale Method

Participants in the test were asked to give their subjective opinion regarding the annoyance they experienced while listening to the noise, on a rating scale from 1 to 10, where 1 represents little annoyance and 10 extreme annoyance. **Table 8** details the corresponding annoyance values reported by the test participants.

Table 8 1-10 rating scale in terms of annoyance.

The analysis of the subjects' responses was carried out with the support of boxplot graphs, which allow for a clear and direct visualization of the distribution of responses, including their mean, median, overall dataset shape, and central tendency.
In **Figure 62**, we can see the distribution of responses from the subjects who participated in the tractor test. The general distribution of values does not show any significant outliers except for the background with a score of 4.
Speed 8 is relatively compact, with a downward trend and a minimum value of 1. Speeds 4 and 12 are more widely distributed, covering almost the entire scale of values.
Only the background shows compactness in the responses, indicating a more uniform impression among participants.
Compared to the average values in the table, it can be said to be approximately representative for all, except for speed 4. However, while the average is representative, it is a limited data point; indeed, with boxplots, we can observe the actual distribution of responses and impressions from the test subjects.

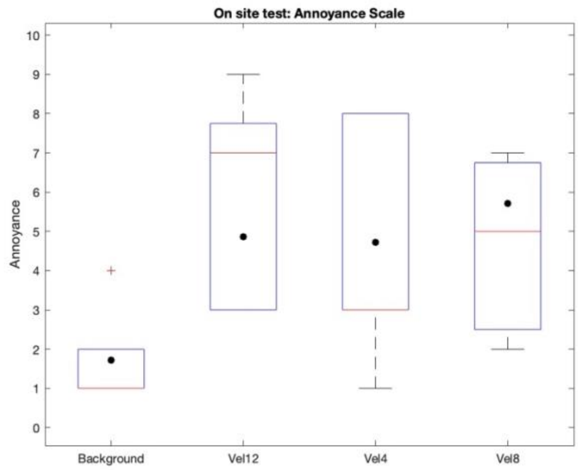


Figure 62 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.1.2. Semantic Differential Method

This section analyzes the results obtained using the Semantic Differential Method as a rating scale. This assessment method evaluates four scales of semantic adjective pairs shown in **Fig. 63**, ranging from -3 to 3 (including 0). The parameters represented on this scale are loudness (represented by Loud/Not Loud), Roughness by Rough/Not Rough, Sharpness by Sharp/Not Sharp, Fluctuation Strength by Fluctuating/Not Fluctuating, and Tonality by Whistling/Not Whistling.

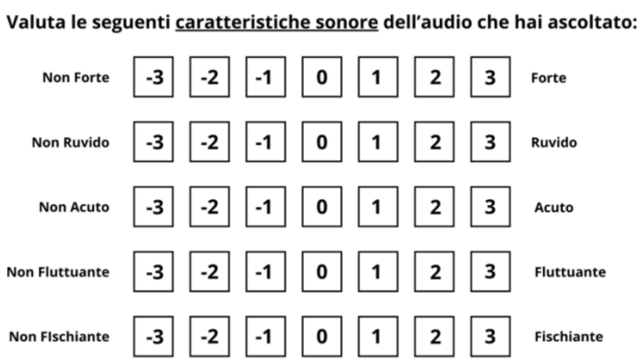


Figure 63 semantic adjective pairs and their relative assessment values; Not Loud(-3)/Loud(3), Not Rough(-3)/Rough(3), Not Sharp(-3)/Sharp(3), Not Fluctuating(-3)/Fluctuating(3), Not Whistling (-3)/Whistling (3).

Figure 64 displays the graph related to responses for the loudness parameter. The background shows low compactness in loudness responses, with responses ranging from -3 to 2.
The averages of the responses are increasing and consistent with the loudness trend.
The distribution of speed 4 is very wide, with values ranging from -3 to 2. Speed 8 has a very limited boxplot and an average close to zero. Speed 12 has a limited boxplot but with values that extend beyond it and a higher average than the others.
Two outliers are present in the responses for speeds 8 and 12, respectively at -3 and -1. These values represent responses that deviate significantly from the general trend.

Compared to the distribution of the objective loudness parameter, we can see that the mean of the responses is consistent, as are the distributions in the boxplots.

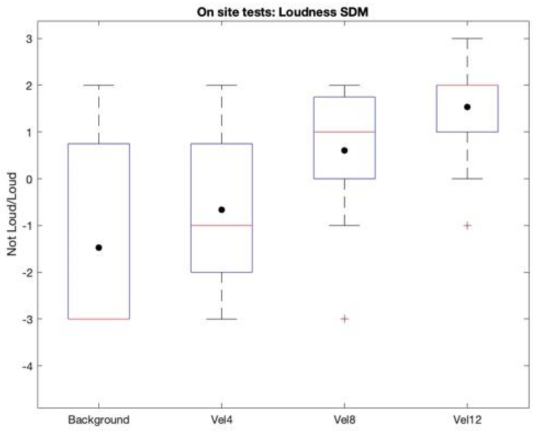


Figure 64 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

The background related to the roughness parameter has a wide range of responses, and the mean and median do not coincide, with the median being higher. Speed 4 has a less wide boxplot, and the mean and median do not coincide, with the mean being higher.
Speed 8 has a narrower range of responses, and the mean coincides with the median. Speed 12 is more compact than speed 8 and has a mean and median that are close, with the mean being lower.
Compared to the objective parameter, the distribution of responses does not consistently follow that of the objective parameter, which has a decreasing trend from speed 4 down to speed 12.
The means and medians also do not follow this trend, as shown in the graph in **Fig. 65.**

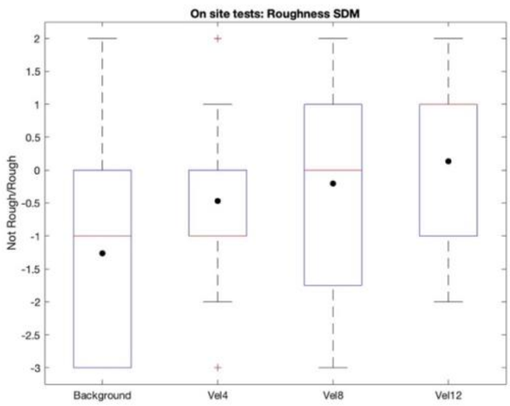


Figure 65 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 66 shows the boxplots related to the sharpness parameter. The background has a wide range of responses, with the mean and median not coinciding, and the mean being higher.
Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being slightly higher.
Speed 8 has a narrower range of responses than those analyzed previously, with the mean slightly higher than the median. Speed 12 is less compact, with a mean and median that are close, with the median being higher.

Compared to the objective parameter, the distribution of responses does not follow the objective parameter’s increasing trend. The means and medians do not follow this trend either.

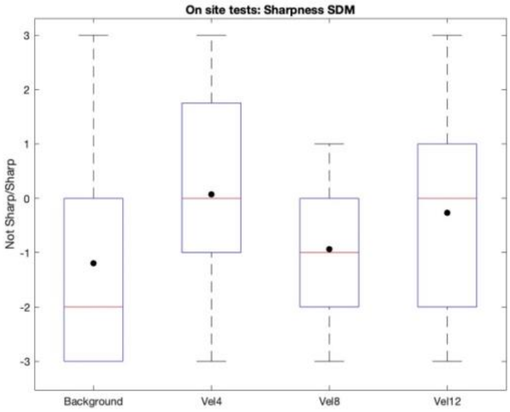


Figure 66 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

The background related to the fluctuation strength parameter has a wide range of responses, with the mean and median not coinciding, and the mean being lower. Speed 4 has a narrower boxplot, and the mean and median do not coincide, with the mean being higher.
Speed 8 has a narrower range of responses than those previously analyzed, with the mean slightly higher than the median. Speed 12 is less compact, with a mean and median that are close, with the median being higher.
Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter’s increasing trend. The means do follow it, with similar values between the background and speed 4 and then increasing. However, the medians do not, as shown in the graph in **Figure 67**.

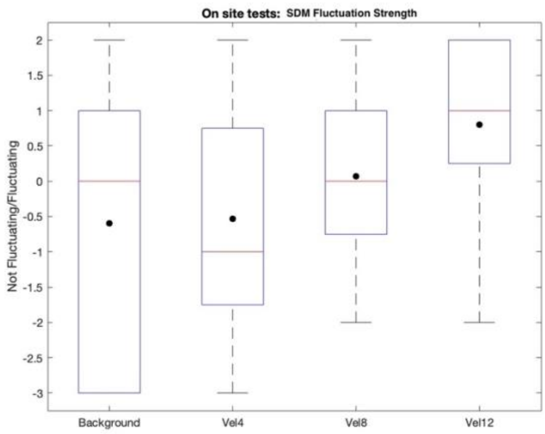


Figure 67 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

In **Figure 68**, we can observe the boxplots related to the tonality parameter. The background has a wide range of responses, and the mean and median do not coincide, with the median being significantly lower.
Speed 4 has a narrower boxplot, with the mean and median not coinciding, showing a small difference and with the mean being lower. Speed 8 has a much narrower range of responses than those previously analyzed, with the mean slightly higher than the median. Speed 12 is not particularly compact, with the mean and median being very close, and the median being slightly higher.

Compared to the objective parameter, the response distribution consistently follows that of the objective parameter, which shows a fluctuating trend. The means and medians also follow this trend.

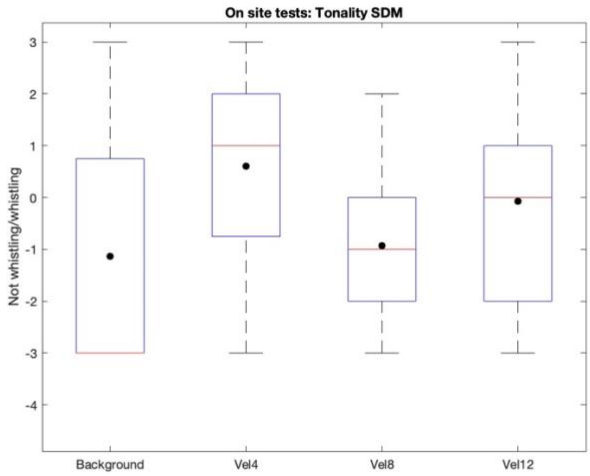


Figure 68 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.1.3. Rating Scale Method: subgroup of acoustics experts and musicians

From the group of subjects examined, only the responses from those who identified as acoustics experts and musicians were selected, in order to evaluate how prior knowledge might influence perception and the responses given. The subgroup counts 9 subjects.
Table 9 shows the average test responses for the subgroup in question.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1.7	-2.4	-1.6	-1.9	-0.9	-1.9
Velocity 4	4.9	-0.4	-1	-0.1	-0.3	0
Velocity 8	4.7	0	-0.9	-1.9	-0.6	-2
Velocity 12	5.7	1.4	-0.3	-1.1	0.7	-1.1

Table 9 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

Figure 69 represents the distribution of responses for the 1-to-10 annoyance rating scale. The background has a compact boxplot at values of 1 and 2, with an outlier at 4. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 shows a broad range of responses, and the mean is much higher than the median. Speed 12 is slightly more compact than speed 8, with the mean and median not close together, and the mean being higher.

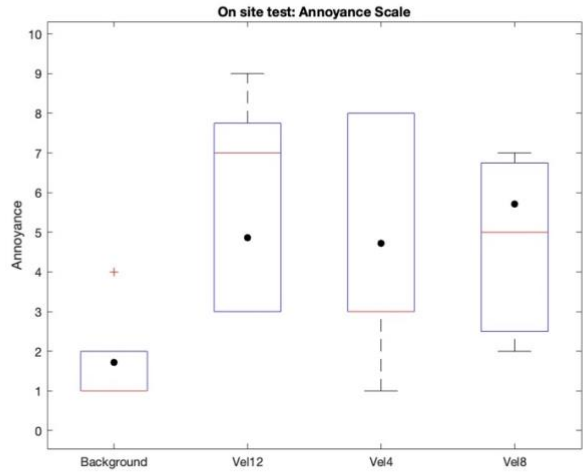
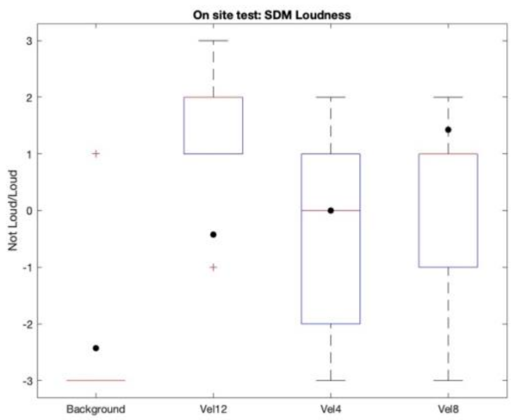
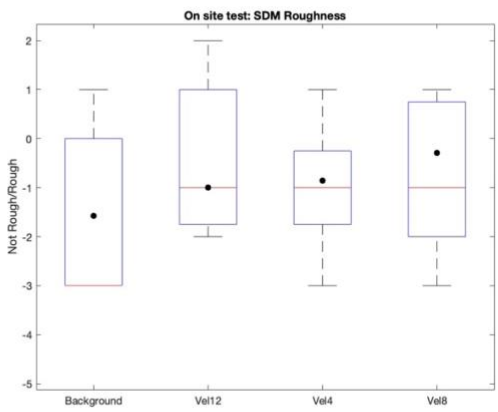


Figure 69 on tractor test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.1.4. Semantic Differential Method: subgroup of acoustics experts and musicians

This section moves on to analyzing the responses of acoustics experts and musicians obtained using the Semantic Differential Method rating scale.
The loudness shown in **Figure 70** did not yield a significant result, as there is no consistency with the objective data, and it differs significantly from the general result. However, the average values follow the trend of the measured loudness.
Roughness aligns with the objective test results, as shown in **Figure 71**. The trend is compact, with an upward tendency in the response distribution. Relative to the objective parameter, there is consistency in response distribution, though not in the averages and medians.
The sharpness trend, as shown in **Figure 72**, does not align with the objective trend for speed 12. For speed 8, however, it is noticeably higher within this subgroup.
The fluctuation strength shown in **Figure 73** is perceived to be relatively high for speed 4. We can observe that it does not follow the objective parameter trend.
Tonality, visible in **Figure 74**, shows consistency with the objective parameters only for speed 4 and the background, while for speeds 8 and 12, the trend is opposite.





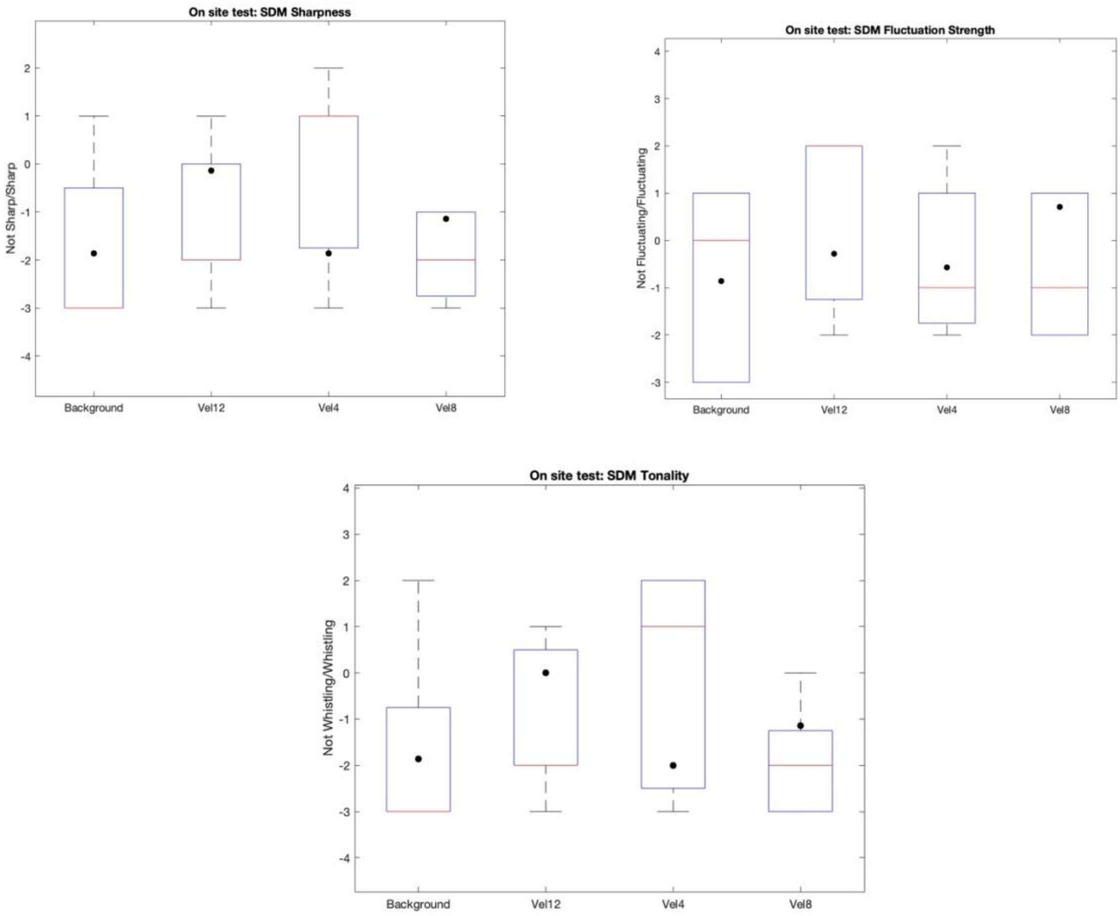


Figure 70-74 on tractor test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.2. Immersive Test

The immersive test involved listening to spatialized audio recordings in ambisonics audio of 3rd order and viewing the interior of the tractor cabin through the use of a VR headset. The operating modes of the HVAC system are recorded with the Zylia ZM-1 microphone and played back to the subjects through a sphere of 16 speakers.

The subjects tested were 47% women and the age varies from 21 years old to 40 years old. All of the subjects are Italian native speaker and two of them have light hearing problems, such as a light form of tinnitus and distortions at very low frequencies.

The results of the subjective ratings are then compared with the psychoacoustic parameters obtained from the NTi omnidirectional microphone. **Table 10** displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1	-3	-3	-3	-2.9	-3
Velocity 4	1.9	-2.1	-1.7	-2.8	-1	-3
Velocity 8	2.3	-0.7	-0.6	-1.4	-0.7	-2.4
Velocity 12	2.9	-0.1	0.5	-1.7	-0.6	-2.1

Table 10 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

3.2.2.1. Rating Scale Method

The following are the results obtained from the analysis of responses based on subjective perceptions measured on the 1-to-10 annoyance rating scale, shown in **Table 8**.
Figure 75 shows the distribution of responses related to annoyance measured with the rating scale method. The background distribution is absolutely compact and unanimous at -3. The other speeds have increasing means (represented by the black dot), and the distribution is more irregular, with speed 4 showing a broad range, speed 8 compacted toward the lower values, and speed 12 compact toward the lower end but with a broader range extending upward.
In comparison with the distribution of Laeq and Lzeq, we can see that the black dots representing the means are consistent, but the boxplot widths do not follow the pattern of the objective parameter.

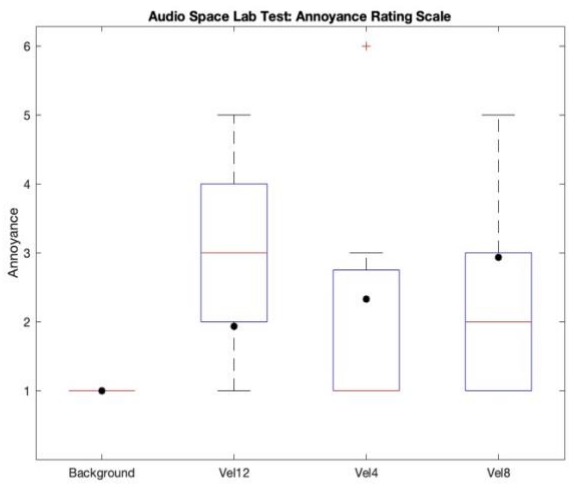


Figure 75 immersive test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.2.2. Semantic Differential Method

Figure 76 shows the distribution for loudness. The background responses are absolutely compact for loudness, with all responses at -3. The average responses increase consistently, in line with the trend of loudness. The distribution for speed 4 is very wide, with values ranging from -3 to 2. Speed 8 has a very limited boxplot and an average around zero. Speed 12 has a limited boxplot, but with values extending beyond it and a higher mean than the others. There are extreme values at both the high and low ends, particularly for speeds 12 and 8.

In comparison to the objective loudness parameter, the mean of the responses is consistent, while the boxplot distributions are not.

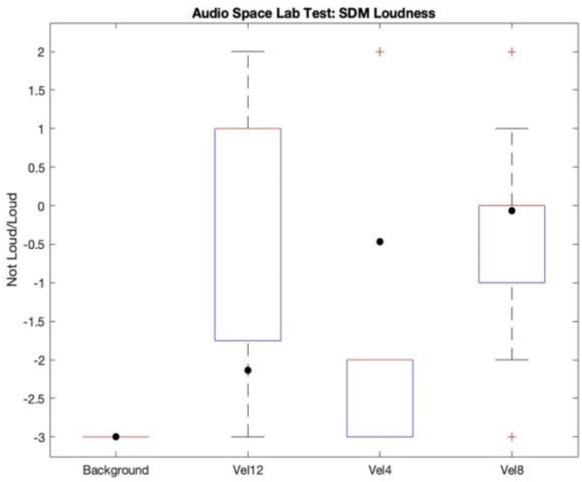


Figure 76 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

For roughness, shown in **Figure 77**, the distribution of background responses is also absolutely consistent and fixed at -3. The means increase significantly from one to the next. For speed 4, we observe an outlier, and the distribution leans toward higher values. Speed 8 has a very wide range, while speed 12 has an outlier and a rather narrow distribution.
In comparison to the objective roughness parameter, the means are not consistent with the distribution recorded by the microphones; however, the boxplots are similar to the objective parameter's distribution, with speed 4 higher than speeds 8 and 12 and the background very low. Speeds 12 and 8 differ slightly from each other and are reversed.

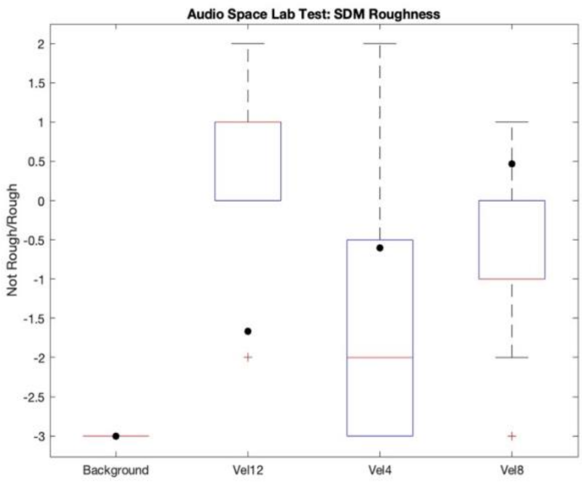


Figure 77 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 78 shows the chart for sharpness. Here again, the background is fixed at -3 with just one outlier at -2. Speed 4 has a mean slightly lower than the median and a much less compact distribution. Speed 8 has a mean

much higher than the median, with two outliers. Speed 12 has a mean higher than the median and a broader distribution, similar to speed 4.
Compared to the objective parameter, there is consistency in terms of distribution, but the mean and median do not follow the trend.

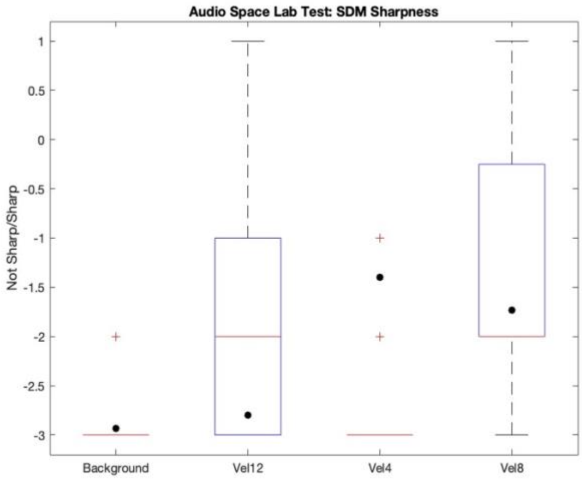


Figure 78 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

The background for fluctuation strength, shown in **Figure 79**, shows no variation, with a compact boxplot at -3, except for one outlier. Speed 4 has a wider boxplot, with the mean and median coinciding. Speed 8 shows a very broad range of responses, with a mean much higher than the median by approximately 1.5. Speed 12 is more compact than speed 8, with the mean and median close together and the mean slightly higher.
Compared to the objective parameter, the response distribution is fairly consistent with that of the objective parameter.

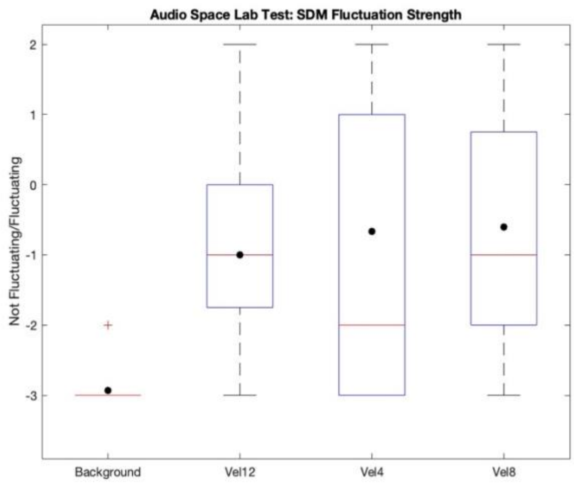


Figure 79 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 80 shows the chart for tonality. The background shows no variation, with a compact boxplot at -3, except for one outlier. Speed 4 has a wider boxplot, with the mean and median coinciding. Speed 8 has a

limited range of responses, with the mean higher than the median. Speed 12 is similar to speed 4, with differing mean and median values, and the mean higher.
In comparison to the objective parameter, the response distribution is fairly consistent with that of the objective parameter, with speed 4 slightly higher than speed 12.

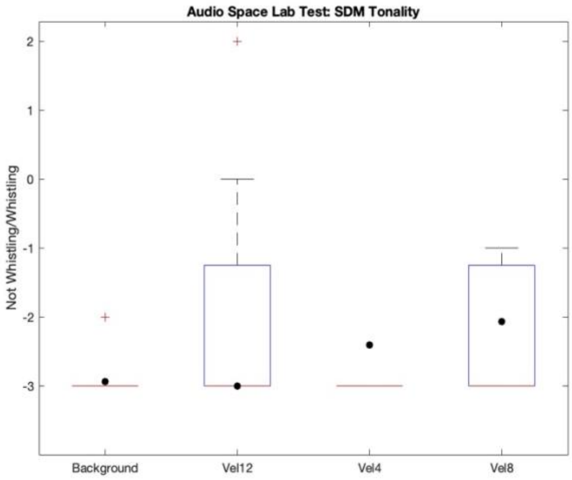


Figure 80 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.2.3. Rating Scale Method: subgroup of acoustics experts and musicians

From the group of subjects examined, only the responses from those who identified as acoustics experts and musicians were selected, in order to evaluate how prior knowledge might influence perception and the responses given. The subgroup counts 10 subjects.
Table 11 displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method, of the subgroup of acoustics experts and musicians.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	3	-1.6	0.2	0	-1.2	1
Velocity 4	3.8	-1.4	1	-1	-1.2	-1.4
Velocity 8	5.2	0.2	2	-0.8	-1.8	-2
Velocity 12	3.8	-0.6	0.8	0	-1.4	-0.8

Table 11 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

Figure 81 displays the distribution of responses related to annoyance. The background shows no variation, with a compact boxplot at -3. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 shows a broad range of responses, and the mean is much higher than the median. Speed 12 is more compact than speed 8, with the mean and median close together, and the mean slightly higher.

Compared to the objective parameter, the response distribution does not consistently follow the trend of the objective parameter, which shows an increasing pattern.

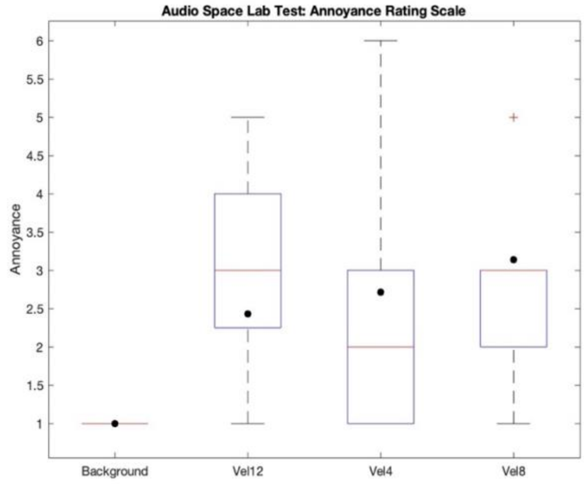


Figure 81 immersive test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

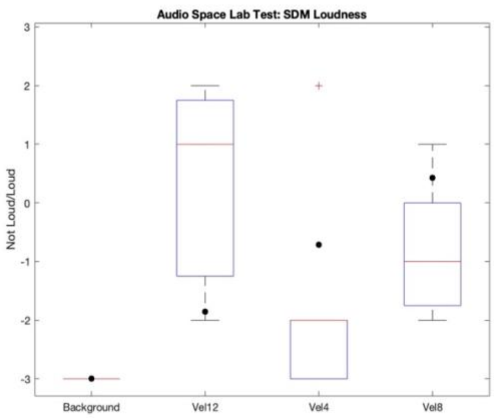
3.2.2.4. Semantic Differential Method: subgroup of acoustics experts and musicians

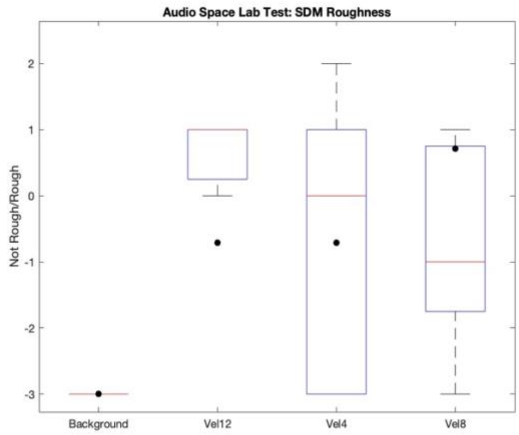
In this paragraph is going to be discussed the distribution of the subjective responses of the : subgroup of acoustics experts and musicians evaluated with the semantic differential method.
In **Figure 82**, we can observe the distribution of responses for loudness assessed using the Semantic Differential Method. The background shows no variation, with a compact boxplot at -3. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 shows a narrow range of responses, with the mean much higher than the median. Speed 12 is less compact than speed 8, with the mean and median not close together, and the mean being higher.
Compared to the objective parameter, the response distribution does not consistently align with the trend of the objective parameter, which shows an increasing pattern. The means follow this trend, but the medians do not.
The background for roughness, shown in **Figure 83**, has no variation, with a compact boxplot at -3. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 covers a broad range of responses, with the mean lower than the median. Speed 12 is more compact than speed 8, with the mean and median not close together, and the mean being higher.
Compared to the objective parameter, the response distribution consistently follows the trend of the objective parameter, which shows a decreasing pattern after speed 4. Both the means and medians follow this trend.
Figure 84 shows the distribution of responses for sharpness. The background shows no variation, with a compact boxplot at -3. Speed 4 has a wider boxplot, with the mean and median coinciding. Speed 8 is compact in terms of response range, with the mean higher than the median. Speed 12 is less compact than speed 8, with the mean and median not close together, and the mean being higher.
Compared to the objective parameter, the response distribution does not consistently align with the trend of the objective parameter, which shows an increasing pattern. The means follow this trend, but the medians do not.
The background for fluctuation strength, represented in **Figure 85**, shows no variation, with a compact boxplot at -3, and an outlier at -2. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being higher. Speed 8 covers a broad range of responses, with the mean much lower than the median. Speed 12 is more compact than speed 8, with the mean and median close together, and the mean being higher.

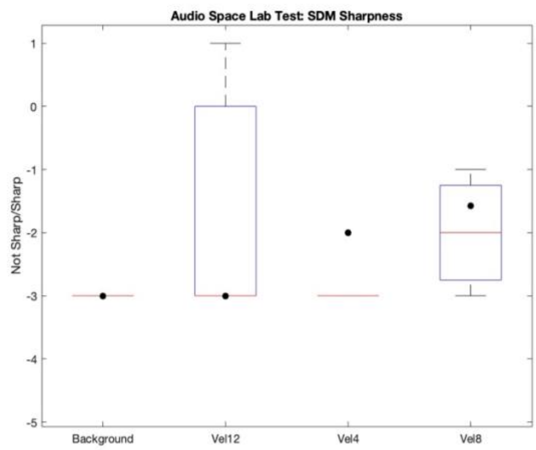
Compared to the objective parameter, the response distribution aligns fairly consistently with the trend of the objective parameter, which shows an increasing pattern, with speed 8 equal to speed 12. In this case, speed 8 is higher than speed 12. Both the means and medians do not follow this trend.

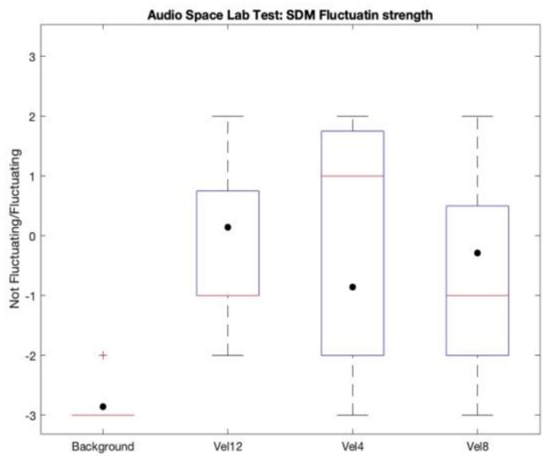
In **Figure 86**, we see the trend of responses for tonality. The background shows no variation, with a compact boxplot at -3. Speed 4 has a wider boxplot, with the mean and median coinciding. Speed 8 shows a narrow range of responses, with the mean slightly higher than the median. Speed 12 is less compact than speed 8, with the mean and median not close together, and the mean being higher. Note the presence of an outlier at -2.

Compared to the objective parameter, the response distribution does not closely follow the trend of the objective parameter, which shows an increasing pattern, with speeds 4 and 8 similar. The means follow this trend approximately, but the medians do not.









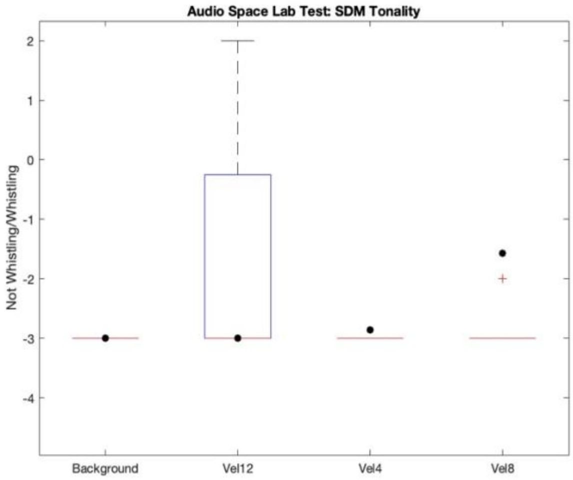


Figure 82-86 immersive test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.3. Immersive Listening Test with Headphones

The test aims to recreate the tractor environment and HVAC sound as accurately as possible using a Meta Quest VR headset and Sennheiser 650 HD headphones. The headset plays a 360° video while the headphones deliver spatialized HVAC audio, achieved through post-recording spatialization, resulting in a third-order ambisonic audio experience.
The subjects tested were 53% men and the age varies from 22 years old to 34 years old. All of the subjects are Italian native speaker and one of them have hearing problems corrected wit hearing aid.
Table 12 displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	2.5	-2.1	0.1	-0.5	-0.4	-0.3
Velocity 4	3.2	-1.5	0.3	-0.8	-0.4	-0.7
Velocity 8	4	-0.1	1.5	-1.5	-1.3	-1.6
Velocity 12	3.2	-0.9	0.5	-0.6	-0.5	-0.7

Table 12 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

3.2.3.1. Rating Scale Method

The following are the results obtained from the analysis of responses based on subjective perceptions measured on the 1-to-10 annoyance rating scale, shown in **Table 8**.
Figure 87 shows the distribution of responses regarding annoyance measured with the 1-10 rating scale mentioned earlier. The background has a lower mean than the median and a relatively narrow distribution. Speed 4 has a wider boxplot, with mean and median almost coinciding. Speed 8 has a response range similar to Speed 4, but the mean is much higher than the median, by approximately 1. Speed 12 is less compact than

Speed 8, with the mean and median close together, though the mean is slightly lower. Compared to the objective parameter, the distribution of responses somewhat consistently follows the objective parameter, with Speeds 4 and 8 being similar but showing a rising trend.

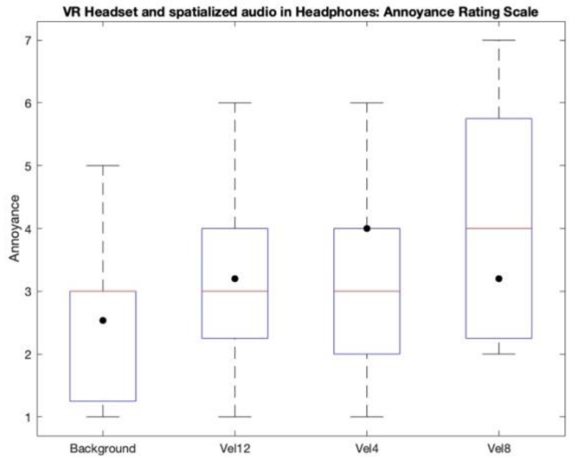


Figure 87 immersive listening test with headphones box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.3.2. Semantic Differential Method

Figure 88 shows the distribution of responses concerning loudness. The background has a lower mean than the median and a relatively narrow distribution, with an outlier at 1. Speed 4 has a wider boxplot, with the mean and median close, though the mean is lower than the median. Speed 8 has a limited range of responses, with the mean about 1 point higher than the median. Speed 12 is relatively compact, with a mean lower than the median. Compared to the objective parameter, the response distribution, aside from Speed 8, aligns with the objective parameter, showing a general upward trend. The medians follow the overall trend, though the means do not.

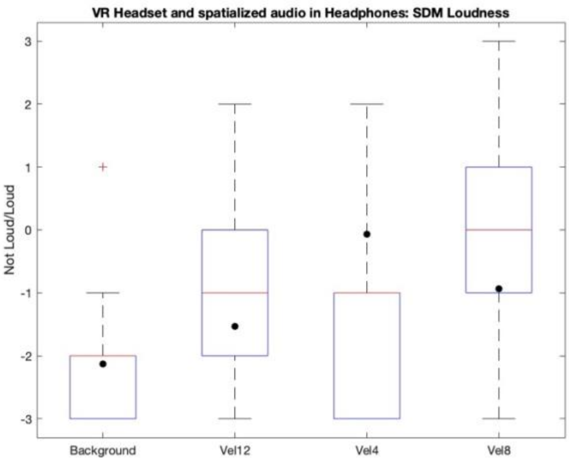


Figure 88 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 89 represents the subjective roughness distribution. The background has a mean very close to the median and a broad distribution. Speed 4 has a narrower boxplot, with the mean slightly below the median. Speed 8 shows a wider response range, with the mean above the median. Speed 12 has a more compact range and a mean lower than the median. Compared to the objective roughness parameter, the distribution is mostly consistent with the objective parameter, with Speed 4 the highest and then decreasing. The medians generally follow this trend, but the means do not.

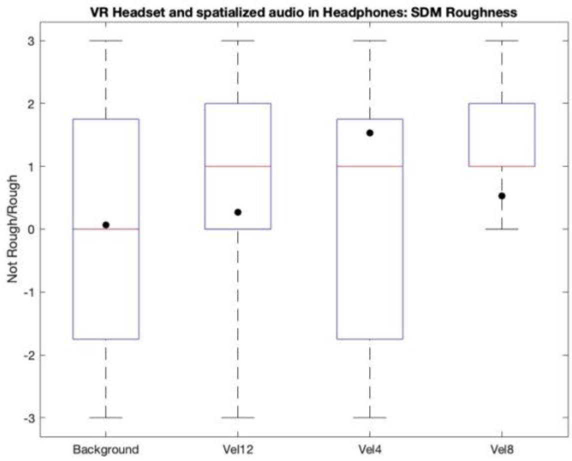


Figure 89 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

The sharpness for the background, shown in **Figure 90**, has a mean slightly above the median and a broad distribution. Speed 4 has a wide boxplot, with the mean and median close, and the mean above the median. Speed 8 has a broad response range, with the mean above the median. Speed 12 is relatively compact, with a mean above the median. Compared to the objective parameter, the response distribution does not align with the objective parameter, which shows an increasing trend with similar values for Speed 4 and background. While background and Speed 4 are indeed similar, the trend decreases with increasing speed, which the median also reflects.

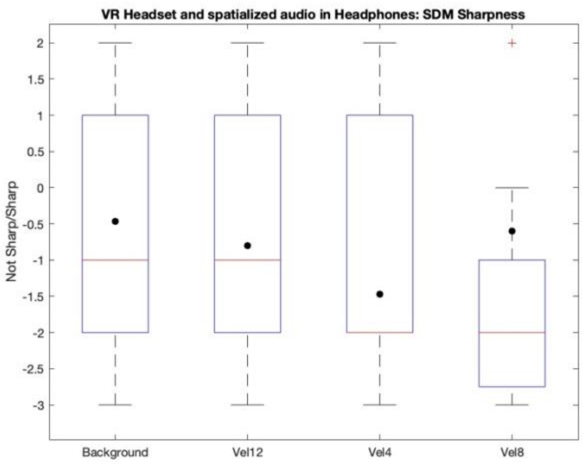


Figure 90 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

In **Figure 91**, we observe the distribution for fluctuation strength. The background has a mean close to, but slightly below, the median and a broad distribution. Speed 4 has a narrower boxplot, with the mean slightly below the median. Speed 8 has a wider response range, with the mean much lower than the median. Speed 12 is relatively compact, with a mean significantly above the median. The distribution of responses does not follow the objective parameter, which has an increasing trend. Both the medians and means do not reflect this trend.

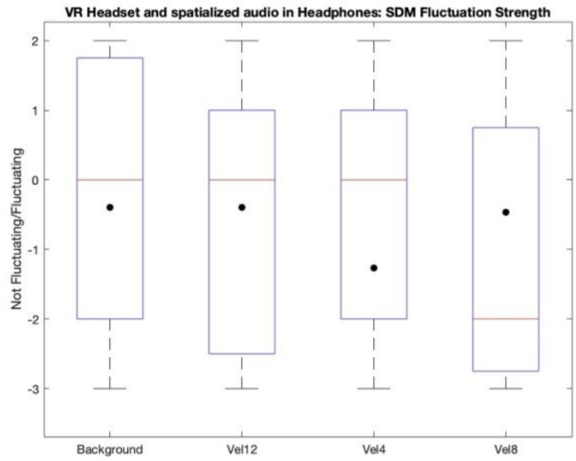


Figure 91 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 92 shows the tonality data. The background has a lower mean than the median and a broad distribution. Speed 4 has a narrower boxplot, with the mean slightly below the median. Speed 8 has a wide range, with the mean below the median. Speed 12 is relatively compact, with a mean much higher than the median. The response distribution does not align with the objective parameter, which shows Speed 4 higher than Speed 8 but lower than Speed 12. Except for the background, the means generally follow this trend, but the medians do not.

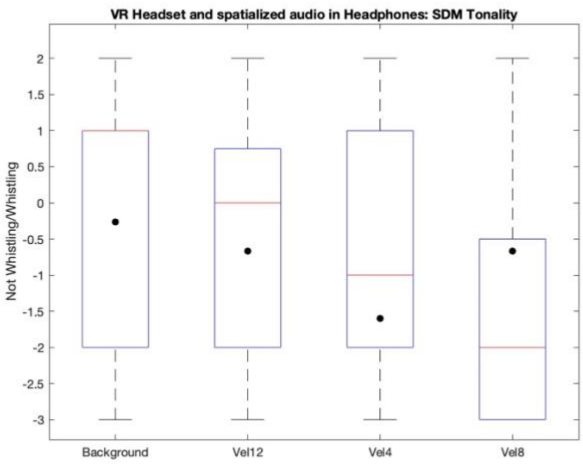


Figure 92 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.3.3. Rating Scale Method: subgroup of acoustics experts and musicians

From the group of subjects examined, only the responses from those who identified as acoustics experts and musicians were selected, in order to evaluate how prior knowledge might influence perception and the responses given. The subgroup counts 5 subjects.
Table 13 displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method, regarding this subgroup.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1	-3	-3	-3	-2.9	-3
Velocity 4	2.4	-1.9	-0.7	-3	0.1	-3
Velocity 8	2.7	-0.7	-0.7	-2	-0.9	-2.9
Velocity 12	3.1	0.4	0.7	-1.6	-0.3	-1.6

Table 11 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

Figure 93 shows the distribution of responses regarding annoyance. The background has a minimal response range, with mean and median coinciding. Speed 4 has a wider boxplot and mean and median do not coincide; the mean is higher. Speed 8 has a relatively broad response range, with the mean higher than the median. Speed 12 is less compact than Speed 8, and its mean and median are quite far apart, with the mean lower. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which shows an increasing trend. Both means and medians follow this trend.

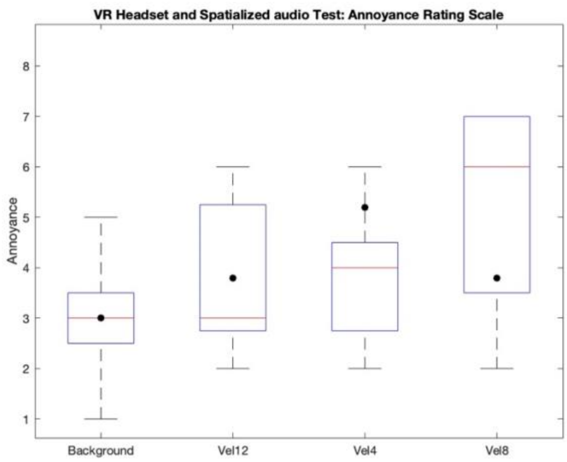


Figure 93 immersive listening test with headphones box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.3.4. Semantic Differential Method: subgroup of acoustics experts and musicians

Figure 94 shows the boxplot for loudness. The background has a minimal response range, with mean and median not coinciding; the mean is higher. Speed 4 has a wider boxplot, with mean and median not coinciding, and the mean is lower. Speed 8 has a relatively broad response range, with the mean higher than the median.

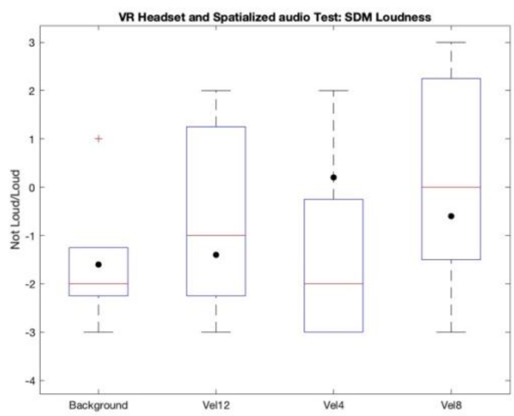
Speed 12 is less compact than Speed 8, and its mean and median are somewhat close, with the mean lower. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which shows an increasing trend. Neither means nor medians follow this trend.

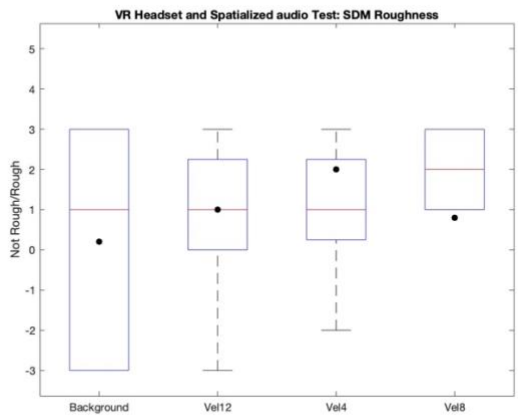
The roughness background, shown in **Figure 95**, has a broad response range, with mean and median not coinciding, and the mean is lower. Speed 4 has a narrower boxplot, with mean and median coinciding. Speed 8 has a narrow response range, with the mean higher than the median. Speed 12 is relatively compact, with mean and median somewhat close, and the mean lower. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which decreases after Speed 4. Neither means nor medians follow this trend.

Figure 96 shows the graph for sharpness. The background has a broad response range, with mean and median not coinciding, and the mean lower. Speed 4 has a wider boxplot, with mean and median not coinciding, and the mean lower. Speed 8 has a broad response range, with the mean higher than the median. Speed 12 is less compact, with mean and median somewhat close, and the mean higher. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which shows an increasing trend. Neither means nor medians follow this trend.

The background shown in **Figure 97**, related to fluctuation strength, shows a broad response range, with mean and median not coinciding, and the mean higher. Speed 4 has a narrower boxplot, with mean and median not coinciding, and the mean lower. Speed 8 has a broader response range, with the mean slightly higher than the median. Speed 12 is more compact than Speed 8, with mean and median somewhat close, and the mean higher. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which shows an increasing trend. Neither means nor medians follow this trend.

Figure 98 shows the graph for tonality. The background has a broad response range, with mean and median not coinciding, and the mean higher. Speed 4 has a narrower boxplot, with mean and median not coinciding, and the mean lower. Speed 8 has a narrower response range, with mean coinciding with the median. Speed 12 is more compact than Speed 8, with mean and median somewhat close, and the mean higher. Compared to the objective parameter, the response distribution does not consistently follow the objective parameter, which shows an increasing trend, with similar values for Speed 8 and Speed 4. Neither means nor medians follow this trend.





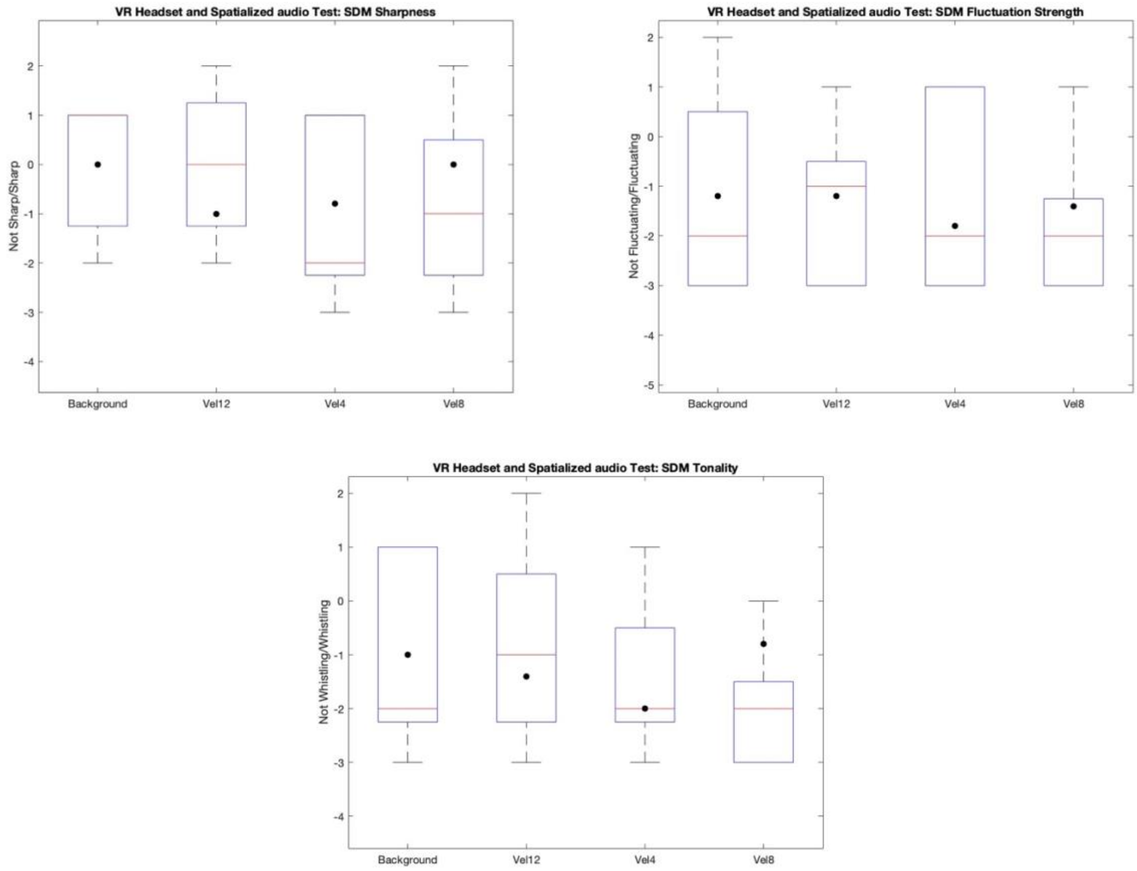


Figure 94-98 immersive listening test with headphones box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.4. Binaural Test

During the test, evaluators listened to binaural recordings made with the artificial head HEAD HSU III.3. The playback was listened with a pair of Sennheiser HD650 headphones, subjects rated the sounds using the 1-10 Annoyance Rating Scale and the Semantic Differential Method. In handling binaural recordings, psychoacoustic parameters are computed individually for each channel. Subsequently, the mean values of these parameters for the left and right channels are determined for each recording across various operational modes.

The subjects tested were 47% women and the age varies from 24 years old to 55 years old. All of the subjects are Italian native speaker, except for two of them having Romanian and Albanian mother tongue. No hearing problems were declared by the subjects.

Table 14 displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1.3	-2.4	-2.4	-2.7	-2.6	-2.8
Velocity 4	1.8	-2	-1.5	-2	-1.9	-2
Velocity 8	3.6	-0.7	-0.6	-1.7	-1.2	-1.9
Velocity 12	5.2	0.5	0.5	-1.2	-0.9	-1.2

Table 14 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

3.2.3.1. Rating Scale Method

The following are the results obtained from the analysis of responses based on subjective perceptions measured on the 1-to-10 annoyance rating scale, shown in **Table 8**.

Figure 99 shows the distribution of responses related to annoyance. The background has a mean very similar to the median, slightly higher, and a very compact distribution. There are two outliers, one at 2 and one at 4. Speed 4 has a narrow boxplot, and the mean and median are quite distant, with the mean being lower than the median. Speed 8 has a narrow range of responses, and the mean is higher than the median. Speed 12 has a broad range of responses and a mean much higher than the median.

Compared to the objective parameter, the distribution of responses does not align with the objective parameter, which has an increasing distribution. The means follow the increasing trend, but the medians do not.

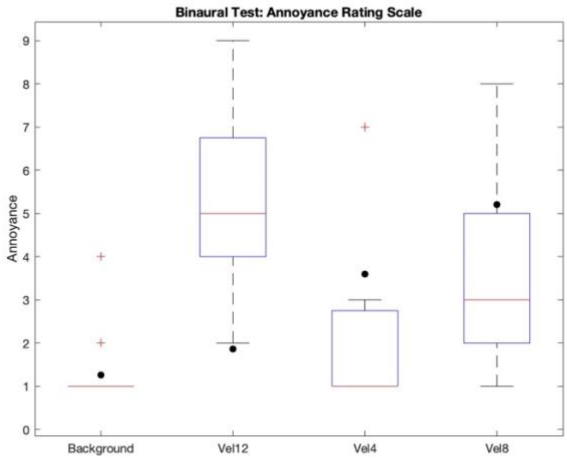


Figure 99 binaural test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.3.2. Semantic Differential Method

Figure 100 shows the distribution of responses related to loudness. The background has a mean similar to the median, slightly higher, and a very compact distribution. There are two outliers. Speed 4 has a very wide boxplot, and the mean and median are quite distant, with the mean being lower than the median. Speed 8 has a narrow range of responses, and the mean is higher than the median. Speed 12 has a broad range of responses and a mean much higher than the median.

Compared to the objective parameter, the distribution of responses does not align with the objective parameter, which has an increasing distribution, due to speed 4. The means follow the increasing trend, but the medians do not.

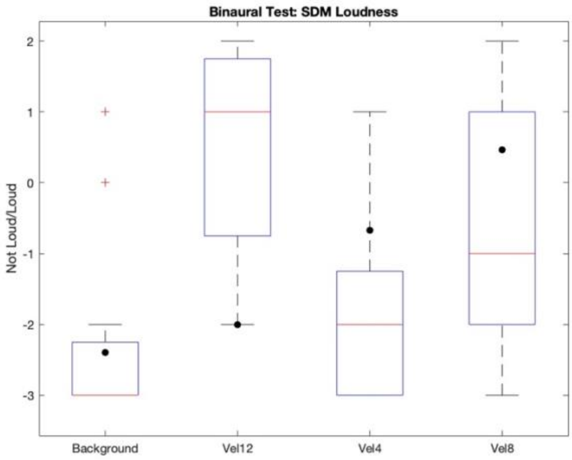


Figure 100 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 101 shows the graph related to roughness. The background has a mean similar to the median, slightly higher, and a very compact distribution. There are two outliers. Speed 4 has a wide boxplot, and the mean and median are quite distant, with the mean being lower than the median. Speed 8 has a narrow range of responses, and the mean is higher than the median. Speed 12 has a broad range of responses and a mean much higher than the median.
Compared to the objective parameter, the distribution of responses is not at all consistent with the objective parameter, which has a decreasing distribution starting from speed 4. Speed 8 is much lower than speed 12, and the means do not follow the distribution. The medians are more similar to the objective parameter.

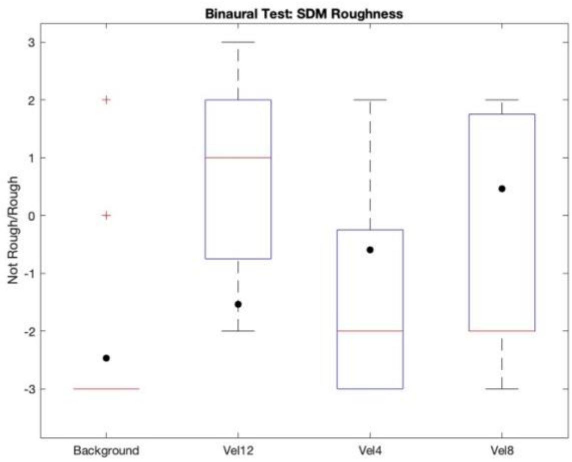


Figure 101 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Sharpness is shown in **Figure 102**. The background has a mean similar to the median, slightly higher, and a very compact distribution. There are two outliers. Speed 4 has a wide boxplot, and the mean and median are

coincident. Speed 8 has a narrow range of responses, and the mean is higher than the median. Speed 12 has a broad range of responses, and the mean is higher than the median.
Compared to the objective parameter, the distribution of responses is not at all consistent with the objective parameter, which has an increasing distribution. The means follow the trend, but the medians do not.

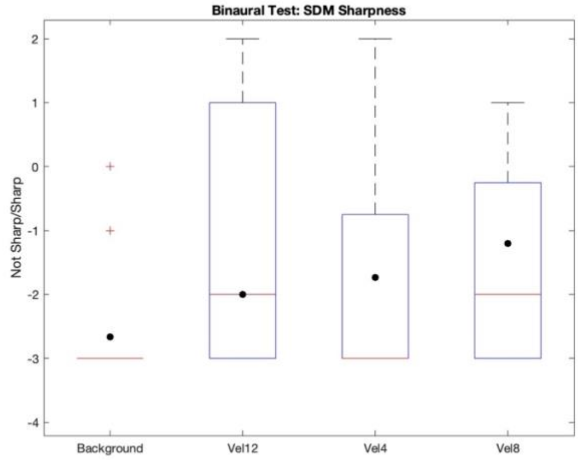


Figure 102 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figure 103 shows the graph related to fluctuation strength. The background has a mean similar to the median, slightly higher, and a very compact distribution. There are three outliers. Speed 4 has a wide boxplot, and the mean and median are not coincident, with the mean being lower. Speed 8 has a wide range of responses, and the mean is much higher than the median. Speed 12 has a broad range of responses and a mean much higher than the median.
Compared to the objective parameter, the distribution of responses is not at all consistent with the objective parameter, which has a higher/similar distribution between speed 4 and 8 and higher for speed 12. The trend is not greatly deviated. The means and medians do not follow the trend.

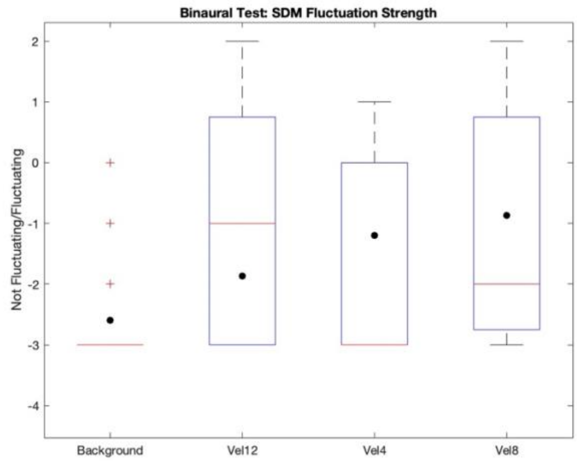


Figure 103 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Tonality is shown in **Figure 104**. The background has a mean similar to the median, slightly higher, and a very compact distribution. There is one outlier. Speed 4 has a wide boxplot, and the mean and median are

coincident. Speed 8 has a narrow range of responses, and the mean is much higher than the median. Speed 12 has a broad range of responses, and the mean is much higher than the median.
Compared to the objective parameter, the distribution of responses is not at all consistent with the objective parameter, which has a higher/similar distribution between speeds 4 and 8 and higher for speed 12. The trend does not deviate much from that of the objective parameter. The means follow the trend, but the medians do not.

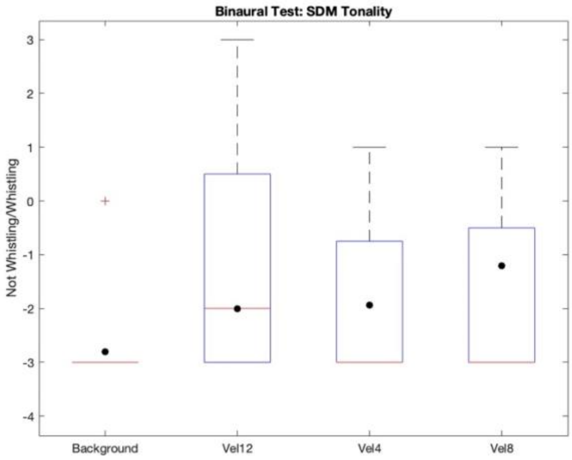


Figure 104 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.4.3. Rating Scale Method: subgroup of acoustics experts and musicians

From the group of subjects examined, only the responses from those who identified as acoustics experts and musicians were selected, in order to evaluate how prior knowledge might influence perception and the responses given. The subgroup counts 9 subjects.
Table 15 displays the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, and the Fluctuation strength, Loudness, Roughness, Tonality and Sharpness scales of semantic differential method, regarding this subgroup.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background	1.6	-1.5	-1.7	-2.2	-2	-2.5
Velocity 4	2.5	-1.5	-1.5	-1.5	-1.7	-1.7
Velocity 8	4.7	0.3	-0.8	-1.7	-0.8	-1.8
Velocity 12	6.2	0.8	0.5	-1.2	-0.8	-1.3

Table 15 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

Figure 105 shows the distribution of responses related to annoyance. The background has a compact range of responses, and the mean and median do not coincide, with the mean being higher. There is one outlier.
Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 has a narrower range of responses, and the mean does not coincide with the median, being higher. There is one outlier. Speed 12 is similarly compact to speed 8 and has the mean and median quite close to each other, with the mean being higher.

Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter, which has an increasing trend. The means do not follow this trend, nor do the medians.

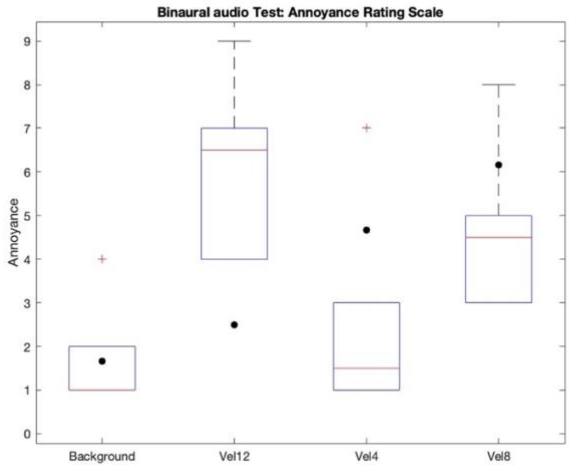


Figure 105 binaural test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

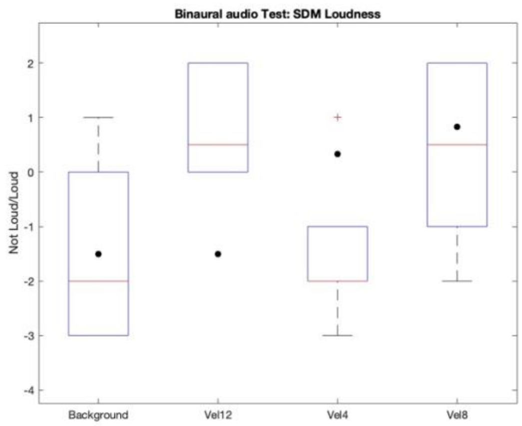
3.2.4.4. Semantic Differential Method: subgroup of acoustics experts and musicians

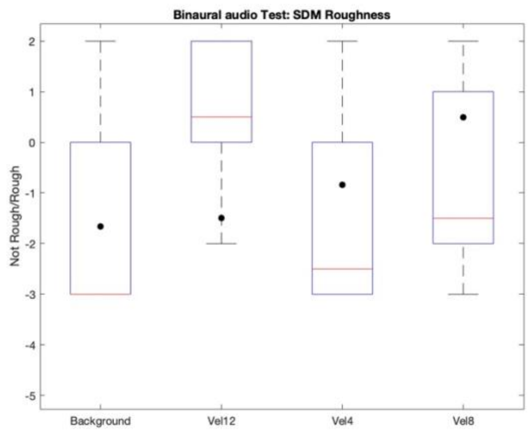
Loudness is shown in **Figure 106**. The background has a less compact range of responses, and the mean and median do not coincide, with the mean being higher. Speed 4 has a narrower boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 has a narrower range of responses, and the mean does not coincide with the median, being higher. There is one outlier. Speed 12 is more compact than speed 8 and has the mean and median quite close to each other, with the mean being higher.
Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter, which has an increasing trend. The means approximately follow it, but the medians do not.
Figure 107 shows the graph related to roughness. The background has a less compact range of responses, and the mean and median do not coincide, with the mean being higher. Speed 4 has a narrower boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 has a broader range of responses, and the mean does not coincide with the median, being higher. Speed 12 is similarly compact to speed 8 and has the mean and median quite close to each other, with the mean being higher.
Compared to the objective parameter, the distribution of responses consistently follows the objective parameter, which has a decreasing trend after speed 4. The means do not follow it, but the medians do.
Sharpness is shown in **Figure 108**. The background has a fairly compact range of responses, and the mean and median do not coincide, with the mean being higher. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being higher. Speed 8 has a broader range of responses, and the mean does not coincide with the median, being higher. Speed 12 is more compact than speed 8 and has the mean and median quite close to each other, with the mean being higher.
Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter, which has an increasing trend. The means do not follow it, nor do the medians.
Figure 109 shows the fluctuation strength. The background has a less compact range of responses, and the mean and median do not coincide, with the mean being higher. Speed 4 has a narrower boxplot, and the mean and median do not coincide, with the mean being lower. Speed 8 has a broader range of responses, and the mean does not coincide with the median, being higher. Speed 12 is similarly compact to speed 8 and has the mean and median quite close to each other, with the mean being higher.

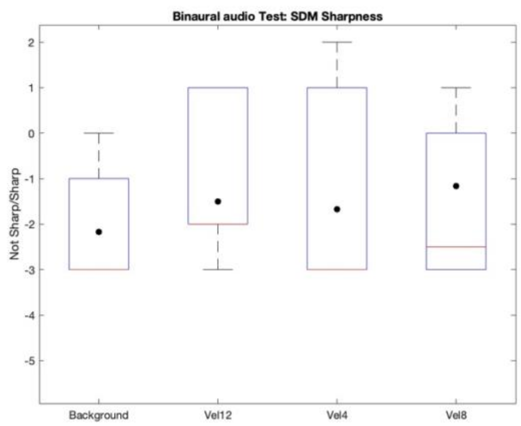
Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter, which has an increasing trend, with speeds 8 and 12 being equal. The means follow it, but the medians do not.

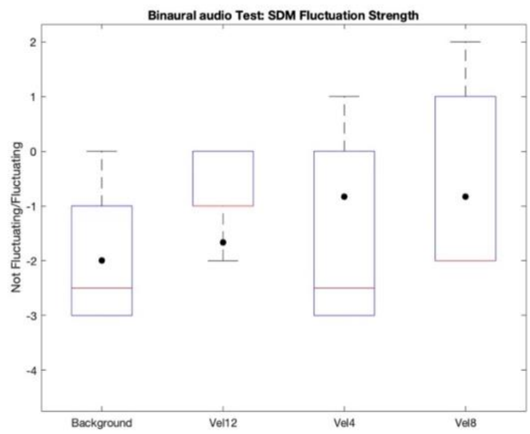
Finally, **Figure 110** shows the graph for tonality. The background has an absolutely compact range of responses, and the mean and median do not coincide, with the mean being higher. Speed 4 has a wider boxplot, and the mean and median do not coincide, with the mean being higher. Speed 8 has a broader range of responses, and the mean does not coincide with the median, being higher. Speed 12 is more compact than speed 8 and has the mean and median quite close to each other, with the mean being higher.

Compared to the objective parameter, the distribution of responses does not consistently follow the objective parameter, which has an increasing trend, with speeds 4 and 8 being similar. The means follow the trend, but the medians do not.









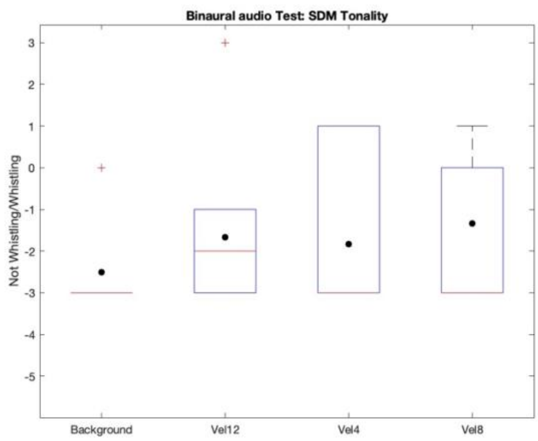


Figure 106-110 binaural test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.5. Compared Test

In order to limit statistical variability and allow for the comparison of the collected data, 5 subjects were selected who completed all three types of laboratory tests mentioned above. The examined subjects are aged between 22 and 40 years. The group consists of 3 women and 2 men. Only one of the five subjects is a musician, and none have hearing issues.

Table 16 reports the mean values of the Annoyance 1-10 Rating Scale for each HVAC mode, as well as the Fluctuation Strength, Loudness, Roughness, Tonality, and Sharpness scales derived from the semantic differential method.

	Annoyance Rating Scale	SDM Loud	SDM Rough	SDM Sharp	SDM Fluctuating	SDM Tonality
Background - *Immersive Test*	1	-3	-3	-3	-3	-3
Velocity 4 - *Immersive Test*	2.6	-1.4	-1.6	-2.8	-0.8	-3
Velocity 8 - *Immersive Test*	2	-0.4	-0.8	-1.8	-0.8	-2.6
Velocity 12 - *Immersive Test*	3.2	-0.4	0.4	-1.6	-0.6	-2
Background - *Binaural Test*	1	-2	-1.4	-1.6	-1.6	-2.4
Velocity 4 - *Binaural Test*	1.4	-0.6	-0.6	-1.2	-2.4	-2.4
Velocity 8 - *Binaural Test*	3.8	1.6	0.8	-0.2	-1.4	-0.6
Velocity 12 - *Binaural Test*	5	-0.6	-0.6	-0.8	-2.4	-1.8
Background - *Immersive VR Test*	2	-3	-3	-3	-3	-3
Velocity 4 - *Immersive VR Test*	2.8	-2.6	-1.4	-2.4	-1.4	-2
Velocity 8 - *Immersive VR Test*	6.6	-0.6	-0.4	-1.8	-0.6	-2.2
Velocity 12 - *Immersive VR Test*	3.2	0.6	0.6	-1.8	-1.2	-1.4

Table 16 For each operating mode, the mean values of the subjective evaluation are provided for the Annoyance Rating Scale and the Semantic Differential Method scales, encompassing Loud/Not Loud, Rough/Not Rough, Sharp/Not Sharp, Whistling/Not Whistling and Fluctuating/Not Fluctuating attributes.

3.2.5.1 Rating Scale Method

The responses related to annoyance, evaluated using the rating scale method in the three tests, are different. The only similarity is the response regarding the background in the headset test and the immersive test. The averages show a similar trend for the binaural test and the immersive test, but only the averages of the responses in the headset test are increasing. An outlier is present in the responses for speed 8 in the binaural test.

Figures 110, 111, and **112** show the distribution of responses for the respective tests.

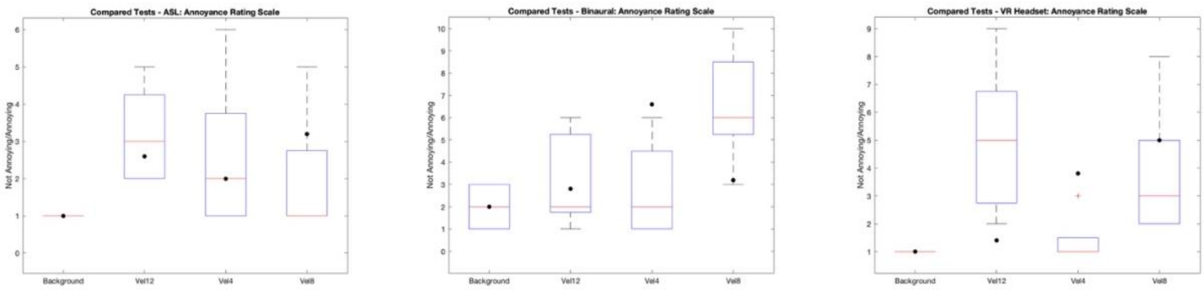


Figure 110-112 compared test box plot for 1-10 rating scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.5.2 Semantic Differential Method

Figures 113, 114, and **115** show the distribution of responses in the comparative test related to loudness. The perception of loudness is neither linear nor consistent across the various tests. While some similarities can be observed in the trend of the averages, the medians, for instance, are not consistent across the tests.
The background in the binaural and immersive tests is identical, whereas the headset test shows a significant difference. An outlier is again present in the binaural test for speed 8.
Averages and medians differ significantly across the various responses.
The objective parameter is increasing, and the only test whose responses align with the increasing trend is the one conducted with the headset.

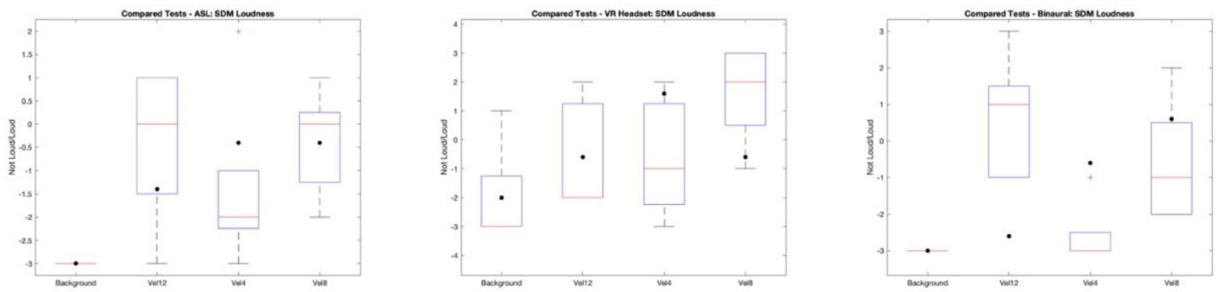


Figure 113-115 compared test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figures 116, 117, and **118** show the distribution of responses in the comparative test related to Roughness. The perceived roughness in the three tests differs significantly, but a certain similarity can be observed in the distributions of the immersive and binaural tests. In contrast, the headset test differs greatly.
The distributions of the immersive and binaural tests follow the trend of the objective parameter, as do their respective medians, though the averages do not. The headset test does not align with the objective parameter, and its distribution is noticeably skewed higher compared to the other tests.

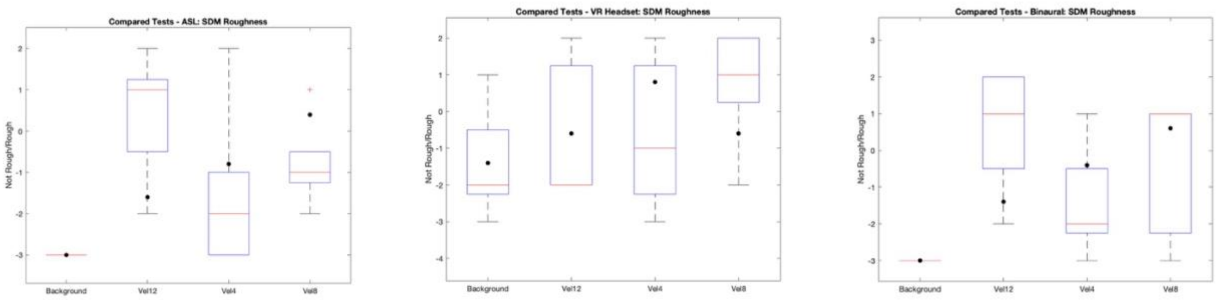


Figure 116-118 compared test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figures 119, 120, and **121** show the distribution of responses in the comparative test related to Sharpness. Sharpness appears similar in the immersive and binaural tests, but these do not follow the increasing trend of the objective parameter in terms of the distribution of responses. However, the averages do respect the increasing trend.
The headset test provides a more accurate representation of the objective parameter, with an increasing trend in both the response distributions and the averages. The medians do not follow the trend, except in the binaural test.
Additionally, it can be observed that the perceived sharpness in the headset test is distributed over higher values compared to the other two tests.

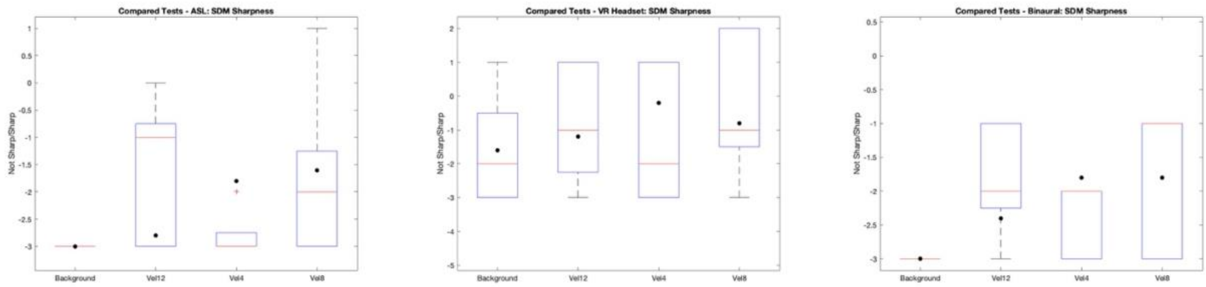


Figure 119-121 compared test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figures 122, 123, and **124** show the distribution of responses in the comparative test related to Fluctuation Strength. Fluctuation strength does not show significant alignment between the tests or with the measured objective parameter, which has an increasing distribution.
The immersive test produced results that are positioned higher on the rating scale compared to the other tests, which are more confined to the lower half.

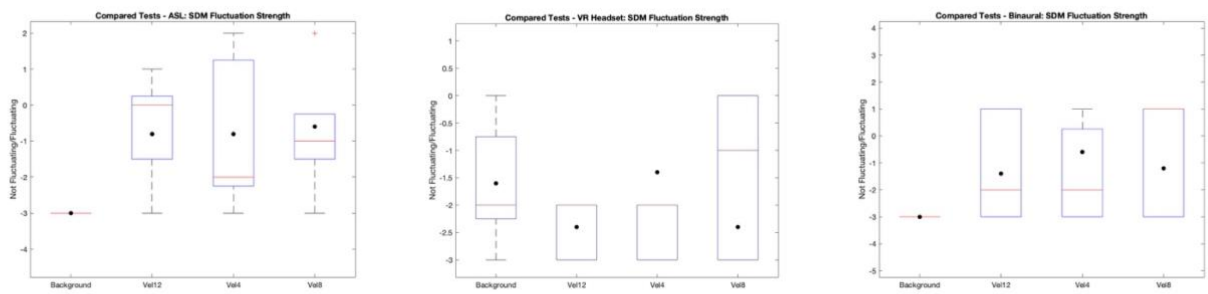


Figure 122-124 compared test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

Figures 125, 126, and **127** show the distribution of responses in the comparative test related to Tonality. Tonality differs significantly across the various tests. In the binaural test, the responses have a wider distribution, whereas in the immersive test, they are much more compact, except for speed 4, where participants noted a more pronounced tonal component. This is also evident in the binaural test responses, where speed 4 shows the widest distribution. The same observation does not apply to the headset test.
In all three tests, however, the responses tend to concentrate in the lower part of the scale. None of the tests show results aligning with the objective parameter, which demonstrates an increasing trend.

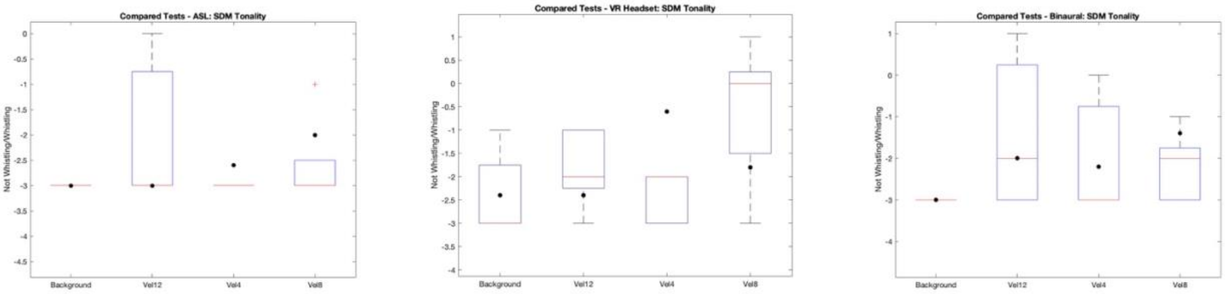


Figure 125-127 compared test box plot for Semantic Differential Method scale; the lower and upper quartile ranges are visually depicted as the bottom and top edges of the blue box, respectively; red line: median; black dot: mean; black T-lines: lowest and highest ratings; '+' symbols: outliers in the data; The HVAC operating modes are shown on the abscissa.

3.2.6 Non-Parametric Correlation Tests: Wilcoxon Tests

In order to assess whether the results obtained, and the data collected from the subjective tests conducted were truly statistically significant, Wilcoxon tests were performed using SPSS software. These are non-parametric tests with related samples, capable of evaluating whether the results collected are due to statistical variability or are indeed representative.
These tests were performed between different pairs of speeds for the same parameter. The significance threshold set was $p<0.005$.

3.2.6.1 On Tractor Test

- The significance level for **annoyance** is good, except for Speed 4 and Speed 8, which would not pass the test even with the threshold raised to 0.1. Therefore, annoyance is consistently higher and increases when comparing lower to higher speeds, except for the Speed 4 and Speed 8 comparison.
- The significance level for **loudness** is good, except for Speed 4 and Background, which would not pass the test even with the threshold raised to 0.1. Therefore, loudness is consistently higher and increases when comparing lower to higher speeds, except for the Speed 4 and Background comparison.
- The significance level for **roughness** is not optimal. Background and Speed 8 would pass the test with the threshold raised to 0.1, but the other tests would not. Therefore, roughness is higher and increases in comparisons of lower to higher speeds only for Speed 4 and Background and for the Speed 12 and Background pair.
- The significance level for **sharpness** is good, except for Speed 4 and Background, which would not pass the test even with the threshold raised to 0.1. Therefore, sharpness is consistently higher and increases when comparing lower to higher speeds, except for Speed 4 and Background.

- The significance level for **fluctuation strength** is acceptable. The combinations that did not pass the test would not pass it even with the threshold raised to 0.1. Therefore, this parameter is higher and increases in comparisons of lower to higher speeds in all comparisons involving Speed 12.
- The significance level for **tonality** is low. The combinations that did not pass the test would not pass it even with the threshold raised to 0.1, except for BG and Speed 4. Therefore, this parameter is higher and increases in comparisons of lower to higher speeds in all comparisons involving Speed 8 and Speed 4, and Speed 8 and Speed 12.

3.2.6.2 Immersive Test

- The significance level for **annoyance** is good, except for Speed 4 and Speed 12, and Speed 12 and Speed 8. The first pair would pass the test with the threshold raised to 0.1, while the second would not. Annoyance is consistently higher and increases when comparing lower to higher speeds, except for the pairs mentioned above.
- The significance level for **loudness** is good, except for Speed 8 and Speed 12, which would not pass the test even with the threshold raised to 0.1. Loudness is consistently higher and increases when comparing lower to higher speeds, except for the pairs mentioned above.
- The significance level for **roughness** is good. Speed 4 and Speed 8 are the only pair that did not pass the test, even if the threshold were raised to 0.1. Roughness is consistently higher and increases when comparing lower to higher speeds, except for the pairs mentioned above.
- The significance level for **sharpness** is good, except for Speed 4 and Background, Speed 8 and Speed 12, which would not pass the test even with the threshold raised to 0.1. Sharpness is consistently higher and increases when comparing lower to higher speeds, except for the pairs mentioned above.
- The significance level for **fluctuation strength** is low. The combinations that did not pass the test would not pass it even with the threshold raised to 0.1. Fluctuation strength is consistently higher and increases when comparing lower to higher speeds only for comparisons involving Background.
- The significance level for **tonality** is good. The combinations that did not pass the test would not pass it even with the threshold raised to 0.1. Tonality is consistently higher and increases when comparing lower to higher speeds, except for the pairs Background and Speed 4, Speed 8 and Speed 12.

3.2.6.3 Immersive Listening Test with Headphones

- The significance level for **annoyance** is good, except for Speed 4 and Background, which would not pass the test even with the threshold raised to 0.1, and for Speed 12 and Speed 4, which also would not pass. Annoyance is consistently higher and increases when comparing lower to higher speeds, except for the pairs mentioned above.
- The significance level for **loudness** is good, except for Speed 8 and Speed 12, which would not pass the test even with the threshold raised to 0.1. Loudness is consistently higher and increases when comparing lower to higher speeds, except for the pair mentioned above.
- The significance level for **roughness** is not optimal. Background and Speed 4, Background and Speed 12, and Speed 4 and Speed 12 would not pass the test even if the threshold were raised to 0.1. Roughness is consistently higher and increases only in comparisons involving Speed 8.
- The significance level for **sharpness** is negligible. Background and Speed 8, and Speed 8 and Speed 12, are the only combinations that would pass the test if the threshold were raised to 0.1. Sharpness never shows a consistent increase when comparing lower to higher speeds.
- The significance level for **fluctuation strength** is low. The combinations that did not pass the test would still not pass even with the threshold raised to 0.1. Fluctuation strength is consistently higher and increases only for the pairs Speed 8 and Speed 12, and Speed 8 and Speed 4 when comparing lower to higher speeds.
- The significance level for **tonality** is low. The combinations that did not pass the test would still not pass even with the threshold raised to 0.1, except for Speed 8 and Speed 4, and Speed 8 and Speed 12.

Tonality is consistently higher and increases only for the pair Speed 8 and Background when comparing lower to higher speeds.

3.2.6.4 Binaural Test

- The significance level for **annoyance** is excellent; all pairs passed the test. Annoyance consistently increases and is higher when comparing lower to higher speeds.
- The significance level for **loudness** is very good, except for Speed 4 and Background, which would not pass the test even with the threshold raised to 0.1. Loudness consistently increases and is higher when comparing lower to higher speeds, except for the pair mentioned above.
- The significance level for **roughness** is very good, except for Speed 4 and Background, which would pass the test if the threshold were raised to 0.1. Roughness consistently increases and is higher when comparing lower to higher speeds, except for the pair mentioned above.
- The significance level for **sharpness** is low. Speed 8 and Background, as well as Speed 12 and Speed 8, would pass the test if the threshold were raised to 0.1, while the other combinations would not. Sharpness consistently increases and is higher only for the pairs Background and Speed 4, and Speed 12 and Speed 4 when comparing lower to higher speeds.
- The significance level for **fluctuation strength** is good. The combinations that did not pass the test would still not pass even with the threshold raised to 0.1. Fluctuation strength consistently increases and is higher when comparing lower to higher speeds, except for the pairs Background and Speed 4, and Speed 12 and Speed 8.
- The significance level for **tonality** is low. The combinations that did not pass the test would do so if the threshold were raised to 0.1, except for Speed 8 and Speed 4. Tonality consistently increases and is higher only for the pairs Background and Speed 12, and Speed 12 and Speed 8 when comparing lower to higher speeds.

3.2.6.5 Compared Test – Immersive Test

- The significance for **annoyance** is very low, except for Speed 12 and Background, which pass the test. The other comparisons would not pass the test even if the threshold were raised to 0.1. Annoyance is higher and shows an increasing trend when comparing lower and higher speeds only for Speed 12 and Background.
- The significance for **loudness** is poor, with only the Background-Speed 8 pair passing the test. Speed 12 and Background, as well as Background and Speed 4, would pass the test if the threshold were raised to 0.1, except for the first pair. Loudness is higher and shows an increasing trend when comparing lower and higher speeds only for the aforementioned pair.
- The significance for **roughness** is low, with only the Background-Speed 8 and Background-Speed 12 pairs passing the test. Roughness is higher and shows an increasing trend when comparing lower and higher speeds only for the aforementioned pairs.
- The significance for **sharpness** is poor, with no pair passing the test. Sharpness is never higher or increasing when comparing lower and higher speeds.
- The significance for **fluctuation strength** is poor, with no pair passing the test. Fluctuation strength is never higher or increasing when comparing lower and higher speeds.
- The significance for **tonality** is poor, with no pair passing the test. Tonality is never higher or increasing when comparing lower and higher speeds.

3.2.6.6 Compared Test – Immersive Listening Test with Headphones

- The significance for **annoyance** is good, except for Speed 12 and Speed 4, Background and Speed 4, and Background and Speed 12, which would not pass the test even if the threshold were raised to 0.1.

Annoyance is always higher and shows an increasing trend when comparing lower and higher speeds, except for the aforementioned pairs.

- The significance for **loudness** is good, except for Speed 12 and Speed 4, Background and Speed 4, and Speed 4 and Speed 8, which would not pass the test even if the threshold were raised to 0.1, except for the last pair. Loudness is always higher and shows an increasing trend when comparing lower and higher speeds, except for the aforementioned pairs.
- The significance for **roughness** is poor, with no pair passing the test. Roughness is never higher or increasing when comparing lower and higher speeds.
- The significance for **sharpness** is poor, with no pair passing the test. Sharpness is never higher or increasing when comparing lower and higher speeds.
- The significance for **fluctuation strength** is poor, with no pair passing the test. Fluctuation strength is never higher or increasing when comparing lower and higher speeds.
- The significance for **tonality** is poor, with no pair passing the test. Tonality is never higher or increasing when comparing lower and higher speeds.

3.2.6.8 Compared Test – Binaural Test

- The significance for **annoyance** is very good, except for Speed 12 and Speed 8, Background and Speed 4, which would not pass the test even if the threshold were raised to 0.1. Annoyance is always higher and shows an increasing trend when comparing lower and higher speeds, except for the aforementioned pairs.
- The significance for **loudness** is good, except for Speed 12 and Speed 8, Background and Speed 4, which would not pass the test even if the threshold were raised to 0.1, except for the first pair. Loudness is always higher and shows an increasing trend when comparing lower and higher speeds, except for the aforementioned pairs.
- The significance for **roughness** is poor, with only the pairs Background-Speed 12 and V4-Speed 12 passing the test. Roughness is higher and shows an increasing trend when comparing lower and higher speeds, but only for the aforementioned pairs.
- The significance for **sharpness** is poor, with no pair passing the test. Sharpness is never higher or increasing when comparing lower and higher speeds.
- The significance **for fluctuation strength** is poor, with no pair passing the test. Fluctuation strength is never higher or increasing when comparing lower and higher speeds.
- The significance for **tonality** is poor, with no pair passing the test. Tonality is never higher or increasing when comparing lower and higher speeds.

3.2.7 Non-Parametric Correlation Tests: Mann-Whitney Tests

In order to evaluate the differences in responses regarding the HVAC speed of the tractor, provided by various participants who took part in different tests, Mann-Whitney tests were conducted. These tests share the same goal as the Wilcoxon tests previously analyzed but are applied to groups consisting of different individuals, whereas the latter involve two groups composed of the same individuals.

The Mann-Whitney tests were carried out by pairing responses obtained from different tests related to the same operating speed being evaluated. The reference threshold for the p-value is 0.05; therefore, the test is considered significant if $p\text{-value} < 0.05$. A significant result indicates that the differences between the two groups being compared are substantial, meaning the distributions of annoyance ratings in the two groups are distinct.

3.2.7.1 Annoyance – Rating Scale

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: Tractor Test-Immersive Test, Compared

Immersive Test-Immersive with Headphones Test, Compared Immersive with Headphones Test-Immersive Test, and Immersive with Headphones Test-Binaural Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of annoyance related to the background noise.

- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: On Tractor Test-Immersive Test, Compared Binaural Test-Immersive with Headphones Test, On Tractor Test-Binaural Test, On Tractor Test-Immersive Test, Immersive Test-Immersive with Headphones Test, and Immersive with Headphones Test-Binaural Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of annoyance related to Speed 4.
- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Tractor Test-Immersive Test, Binaural Test-Compared Immersive Test-Immersive with Headphones Test, Immersive Test-Immersive with Headphones Test, and Compared Immersive with Headphones Test-Immersive Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of annoyance related to Speed 8.
- Regarding **Speed 12**, the pairs that resulted in values below the threshold are: On Tractor Test-Immersive Test, Binaural Test-Immersive with Headphones Test, Immersive Test-Binaural Test, Compared Immersive Test-Binaural Test, and On Tractor Test-Compared Immersive Test. Thus, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of annoyance related to Speed 12.

3.2.7.2 Loudness – Semantic Differential Method

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: Tractor Test-Immersive Test, Compared Immersive Test-Immersive with Headphones Test, Compared Immersive with Headphones Test-Immersive Test, Compared Immersive Test-Binaural Test, Immersive Test-Binaural Test, and Immersive with Headphones Test-Immersive Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of loudness related to the background.
- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: On Tractor Test-Immersive Test, On Tractor Test-Binaural Test, and Compared Binaural Test-On Tractor Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of loudness related to Speed 4.
- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Tractor Test-Immersive Test, Compared Immersive with Headphones Test-Binaural Test, and Compared Immersive with Headphones Test-Immersive Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of loudness related to Speed 8.
- Regarding **Speed 12**, the pairs that resulted in values below the threshold are: On Tractor Test-Immersive Test, Binaural Test-Immersive with Headphones Test, On Tractor Test-Compared Immersive Test, and Compared Immersive with Headphones Test-On Tractor Test. Thus, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of loudness related to Speed 12.

3.2.7.3 Roughness – Semantic Differential Method

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: On Tractor Test-Immersive Test, Compared Immersive Test-Immersive with Headphones Test, Compared Immersive with Headphones Test-Immersive Test, Compared Immersive Test-Binaural Test, On Tractor Test-Compared Immersive Test, Immersive with Headphones Test-Binaural Test, Immersive with Headphones Test-Compared Binaural Test, and Compared Binaural Test-On Tractor Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of roughness related to the background.

- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: On Tractor Test-Immersive Test, Immersive with Headphones Test-Immersive Test, and Immersive with Headphones Test-Binaural Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of roughness related to Speed 4.
- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Immersive with Headphones Test-Binaural Test, Immersive with Headphones Test-Immersive Test, Immersive with Headphones Test-Compared Binaural Test, Tractor Test-Immersive with Headphones Test, and Compared Immersive Test-Immersive with Headphones Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of roughness related to Speed 8.
- Regarding **Speed 12**, no pair showed a result below the threshold. Thus, the results of the tests for all pairs revealed similar distributions. This indicates that participants had the same perception of roughness related to Speed 12.

3.2.7.4 Sharpness – Semantic Differential Method

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: On Tractor Test-Immersive Test, Compared Binaural Test-On Tractor Test, Binaural Test-On Tractor Test, Immersive Test-Immersive with Headphones Test, Immersive with Headphones Test-Binaural Test, On Tractor Test-Compared Immersive Test, Compared Immersive with Headphones Test-Immersive Test, Immersive with Headphones Test-Compared Immersive Test, and Immersive with Headphones Test-Compared Binaural Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of sharpness related to the background.
- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: Immersive with Headphones Test-Binaural Test, Immersive with Headphones Test-Immersive Test, On Tractor Test-Immersive Test, On Tractor Test-Compared Immersive Test, Immersive with Headphones Test-Compared Immersive Test, Compared Immersive with Headphones Test-Immersive Test, and Compared Binaural Test-On Tractor Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of sharpness related to Speed 4.
- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Immersive with Headphones Test-Binaural Test, Immersive with Headphones Test-Immersive Test, On Tractor Test-Immersive Test, Compared Immersive with Headphones Test-Binaural Test, Immersive with Headphones Test-Compared Immersive Test, and Compared Binaural Test-Immersive Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of sharpness related to Speed 8.
- Regarding **Speed 12**, only one pair resulted in a value below the threshold: On Tractor Test-Immersive Test. Therefore, the results of the tests for all pairs revealed similar distributions, with the exception of the aforementioned pair. This indicates that participants had the same perception of sharpness related to Speed 12.

3.2.7.5 Fluctuation Strength – Semantic Differential Method

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: On Tractor Test-Immersive Test, Immersive Test-Immersive with Headphones Test, Immersive Test-Binaural Test, Immersive with Headphones Test-Compared Binaural Test, Binaural Test-Compared Binaural Test, Binaural Test-Compared Immersive Test, Immersive with Headphones Test-Compared Immersive Test, and Compared Immersive with Headphones Test-Immersive Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of fluctuation strength related to the background.
- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: Immersive with Headphones Test-Binaural Test, Binaural Test-On Tractor Test, Immersive with Headphones Test-Compared Immersive with Headphones Test, and Compared Immersive with Headphones Test-On Tractor Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of fluctuation strength related to Speed 4.
- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Immersive with Headphones Test-On Tractor Test, Compared Immersive with Headphones Test-On Tractor Test, and On Tractor Test-Binaural Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of fluctuation strength related to Speed 8.
- Regarding **Speed 12**, the pairs that resulted in values below the threshold are: On Tractor Test-Binaural Test, Immersive with Headphones Test-On Tractor Test, Immersive Test-On Tractor Test, Compared Immersive with Headphones Test-On Tractor Test, Compared Immersive Test-On Tractor Test, On Tractor Test-Compared Binaural Test, and Compared Immersive with Headphones Test-Immersive with Headphones Test. Thus, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of fluctuation strength related to Speed 12.

3.2.7.6 Tonality – Semantic Differential Method

- The comparisons performed for the values obtained from responses related to the **Background** showed a result below the threshold for the following pairs: On Tractor Test-Immersive Test, Immersive Test-Immersive with Headphones Test, Immersive Test-Binaural Test, Immersive with Headphones Test-Binaural Test, Binaural Test-On Tractor Test, Immersive with Headphones Test-Compared Binaural Test, Immersive with Headphones Test-Compared Immersive Test, and Immersive with Headphones Test-Compared Immersive with Headphones Test. Therefore, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of tonality related to the background.
- Regarding **Speed 4**, the pairs that resulted in values below the threshold are: Immersive with Headphones Test-Binaural Test, Compared Binaural Test-On Tractor Test, Immersive Test-Binaural Test, Immersive with Headphones Test-Immersive Test, Immersive Test-On Tractor Test, Binaural Test-On Tractor Test, Immersive Test-Compared Binaural Test, Compared Immersive Test-On Tractor Test, Immersive with Headphones Test-Compared Immersive Test, Compared Immersive with Headphones Test-Immersive Test, and Compared Immersive with Headphones Test-On Tractor Test. Thus, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of tonality related to Speed 4.

- The comparisons performed for the values obtained from responses related to **Speed 8** showed a result below the threshold for the following pairs: Compared Immersive with Headphones Test-Immersive Test, Compared Immersive Test-On Tractor Test, and Immersive Test-On Tractor Test. Therefore, the results of the tests for the mentioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of tonality related to Speed 8.
- Regarding **Speed 12**, the pairs that resulted in values below the threshold are: Immersive Test-On Tractor Test, On Tractor Test-Binaural Test, Immersive with Headphones Test-Immersive Test, and Immersive with Headphones Test-Binaural Test. Thus, the results of the tests for the aforementioned pairs revealed distributions that are significantly different. This indicates that participants did not have the same perception of tonality related to Speed 12.

3.3 Prediction Model

Multiple linear regression analyses were conducted to explore the relationship between subjective assessments and psychoacoustic parameters for both the immersive test and the binaural listening test. Initially, a correlation analysis between subjective and objective data was conducted to identify the psychoacoustic parameters most relevant for predicting subjective ratings. Selection of independent variables for the prediction model was based on their correlation with subjective ratings and consideration of multicollinearity among the psychoacoustic parameters. High correlation between independent variables could compromise the reliability of the model. Therefore, variables with low multicollinearity were preferred. Additionally, significance tests previously analyzed such as Wilcoxon tests and Withney tests were considered to ensure the reliability of the prediction model.

3.3.1 On Tractor Test

Tables 17, 18 display all the data used for the prediction model. The independent variables employed for multiple linear regression are the psychoacoustic parameters computed from the NTi omnidirectional microphone recorded samples and from the artificial head HSU III.3 from Head Acoustics, in order to compare the difference in these terms of the microphones. The psychoacoustic parameters extracted from the artificial head are reported as the average between the left and right channels for each recording. The subjective ratings will be regarded as dependent variables. If an off-diagonal element of P is smaller than the significance level (default is 0.05), then the corresponding correlation in R is considered significant.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluct. Str.	Annoyance Scale	SDM Fluctuation Strength	SDM Loudness	SDM Tonality	SDM Roughness	SDM Sharpness
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	1.8	-0.6	-1.5	-1.1	-1.3	-1.2
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	4.2	-0.5	-0.7	0.2	-0.5	0
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	4.9	0.1	0.6	-0.9	-0.2	-0.9
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	6	0.8	1.5	-0.1	0.1	-0.3

Table 17 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluct. Str.	Annoyance Scale	SDM Fluctuation Strength	SDM Loudness	SDM Tonality	SDM Roughness	SDM Sharpness
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	1.8	-0.6	-1.5	-1.1	-1.3	-1.2
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	4.2	-0.5	-0.7	0.2	-0.5	0
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	4.9	0.1	0.6	-0.9	-0.2	-0.9
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	6	0.8	1.5	-0.1	0.1	-0.3

Table 18 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with omnidirectional microphone NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.1.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 19**. In **Table 20** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Fluctuation Strength	SDM Loudness	SDM Tonality	SDM Roughness	SDM Sharpness
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.99307259	0.927301255	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.55091071	0.718642653	0.45897171	0.471573312	1								
Tonality	0.94444507	0.862469629	0.95647201	0.886317821	0.507296149	1							
Fluctuation Strength	0.9262928	0.969661811	0.88882791	0.953779286	0.610116675	0.755660283	1						
Annoyance Scale	0.92215193	0.973681862	0.87594562	0.873536053	0.830467371	0.871129537	0.893609926	1					
SDM Fluctuation Strength	0.98989838	0.92309332	0.99805924	0.988298571	0.427024734	0.937420451	0.899646763	0.858077638	1				
SDM Loudness	0.990803	0.989237128	0.96965198	0.985074729	0.615296009	0.899433738	0.965851065	0.943799015	0.968896366	1			
SDM Tonality	**0.33549714**	**0.366689281**	**0.30355663**	**0.177429577**	**0.706191671**	**0.538607187**	**0.133533572**	0.564125338	0.244196011	0.309168079	1		
SDM Rough	0.91854963	0.979171833	0.86884951	0.876827927	0.834727382	0.850232653	0.910082283	0.998564535	0.853401578	0.946500302	0.530267283	1	
SDM Sharp	**0.35522646**	**0.410200452**	**0.31259617**	**0.20173174**	**0.770624147**	**0.532191099**	**0.185659445**	0.605725863	0.254558441	0.34187285	0.995062758	0.576154513	1

Table 19 Correlation coefficients for the tractor test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness,

Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales. The bold numbers don't make the assumption (p-value<0.05).

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Fluctuation Strength	SDM Loudness	SDM Tonality	SDM Roughness	SDM Sharpness
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.96061386	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.95816669	0.84688846	0.83220306	1								
Tonality	0.98274366	0.9379216	0.9577757	0.91049069	0.87514514	1							
Fluctuation Strength	0.94064849	0.98089541	0.95354956	0.97454692	0.90837314	0.8749913	1						
Annoyance Scale	0.97529582	0.97977799	0.90132029	0.87999909	0.99196134	0.92871657	0.92869818	1					
SDM Fluctuation Strength	0.94576301	0.92828598	0.99499341	0.98164875	0.79257585	0.9503895	0.92308287	0.85807764	1				
SDM Loudness	0.97900056	0.98960663	0.98865668	0.98748982	0.90717709	0.94375835	0.98466617	0.94379902	0.96889637	1			
SDM Tonality	0.49333802	0.40178745	0.28125593	0.17051541	0.57567439	0.53372317	0.22002125	0.56412534	0.24419601	0.30916808	1		
SDM Rough	0.96918074	0.98279149	0.89913336	0.88513099	0.99420805	0.9137171	0.9404757	0.99856454	0.85340158	0.9465003	0.53026728	1	
SDM Sharp	0.51876788	0.44204612	0.30030863	0.19866125	0.62670176	0.54103072	0.26751953	0.60572586	0.25455844	0.34187285	0.99506276	0.57615451	1

Table 20 Correlation coefficients for the tractor test. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales. The bold numbers don't make the assumption (p-value<0.05).

Starting with the data reported in **table 19**, the analysis revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.97, followed by LZeq at 0.92. In contrast, Roughness exhibited the lowest correlation at 0.81, therefore roughness was omitted from the model due to its low correlation with subjective annoyance. LZeq is not used due to its collinearity with LAeq. After some tests, the best prediction model for noise annoyance relied on LAeq, resulting in an **adjusted R-square** value of 0.92 and a Root Mean Square Error (**RMSE**) of 0.49. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (10) has been determined. **Figure 128** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -0.633 + 0.098 \times LAeq \tag{10}$$

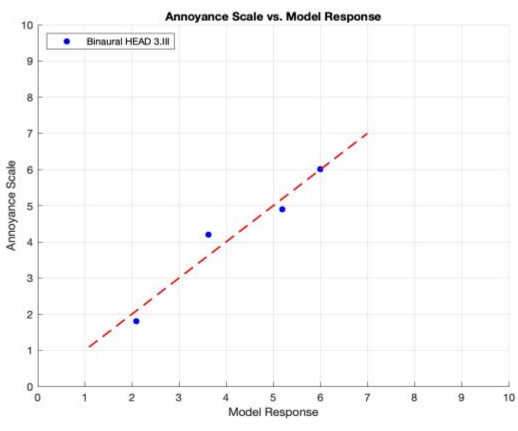


Fig 128 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (10). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 20**, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.97, followed by LAeq at 0.94. In contrast, Tonality exhibited the lowest correlation at 0.66, therefore tonality was omitted from the model due to its low correlation with subjective annoyance. After some tests, the best prediction model for noise annoyance relied on Roughness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.04. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (11) has been determined. **Figure 129** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -3.722 + 26.921 \times Roughness + 0.66 \times LZeq \quad (11)$$

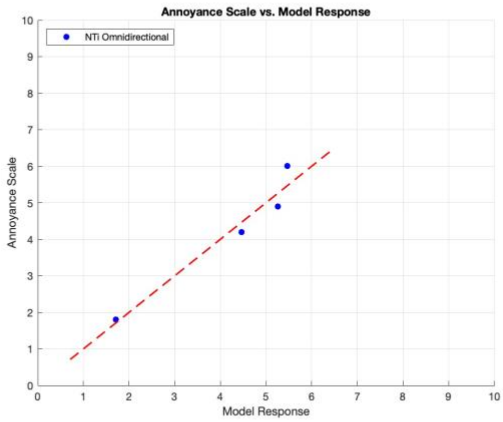


Fig 129 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (11). The straight red line indicates the cases where model responses are equal to subjective responses.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.

Table 21 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 22** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	1.8
V4	33.5	32.5	35	32.5	4.2
V8	50	50	52	47.5	4.9
V12	59	60	62	57	6

Table 21 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99977304	1			
PNC	0.99815903	0.999209045	1		
RNC	0.99660463	0.994984048	0.99037924	1	
Annoyance Scale	0.96767082	0.972747591	0.98038742	0.950236454	1

Table 22 Correlation coefficients for the tractor test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Preferred Noise Criterion, resulting in an **adjusted R-square** value of 0.94 and a Root Mean Square Error (**RMSE**) of 0.43. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (12) has been determined. **Figure 130** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are better than the ones reported above, showing that the parameters based on spectral shape are a better indicator of the annoyance prediction model.

$$Annoyance = 0.710 + 0.866 \times PNC \qquad (12)$$

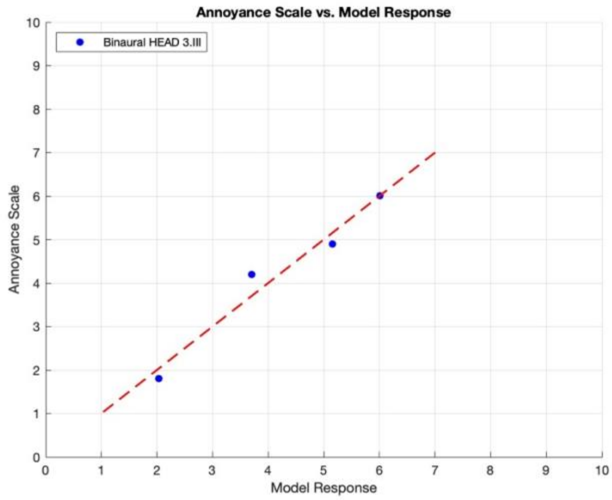


Fig 130 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (12). The straight red line indicates the cases where model responses are equal to subjective responses

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 23** shows the data used for the prediction and **table 24** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	1.8
V4	25	24	23	25	4.2
V8	42	41	42	40	4.9
V12	50	52	51	48	6

Table 23 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.998357164	1			
PNC	0.999999008	0.998367538	1		
RNC	0.999750838	0.999383483	0.99975183	1	
Annoyance Scale	0.971242516	0.970395992	0.971575676	0.970877139	1

Table 24 Correlation coefficients for the tractor test. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Noise Criterion and Noise Rating, resulting in an **adjusted R-square** value of 0.998 and a Root Mean Square Error (**RMSE**) of 0.06. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (13) has been determined. **Figure 131** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are better than the ones reported above, showing that the parameters based on spectral shape are a better indicator of the annoyance prediction model.

$$Annoyance = 75.42 - 16.3 \times NR + 14.62 \times NC \quad (13)$$

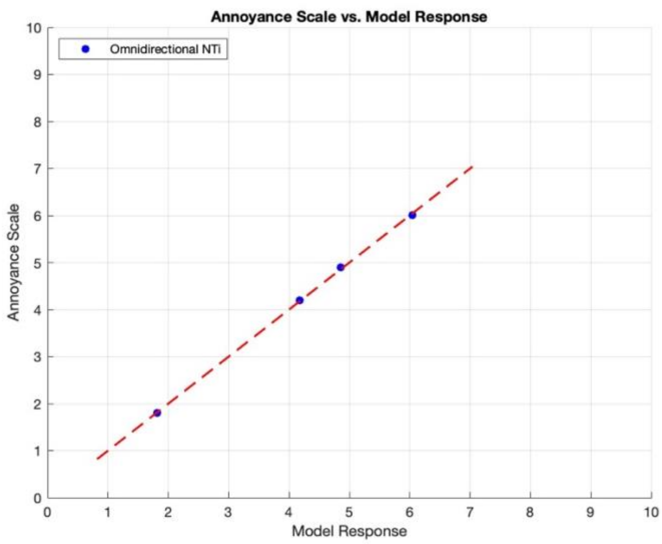


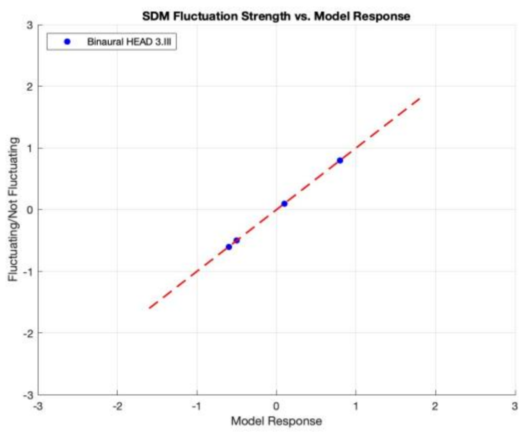
Fig 131 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (13). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.1.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 19** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 20** the ones from the omnidirectional microphone NTi. **Table 21** and **22** reports the correlation coefficients used.
Starting with the data reported in **table 19**, the analysis for the subjective parameter **Fluctuation Strength** revealed a hierarchy of correlation strength, with Loudness demonstrating the highest correlation at 0.989, followed by LZeq at 0.987. After some tests, the best prediction model for Fluctuation Strength relied on LZeq and Roughness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.001. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (14) has been determined.
The analysis of the correlation reported in **table 20**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation strength, with Loudness demonstrating the highest correlation at 0.998, followed by LZeq at 0.997. After some tests, the best prediction model for Fluctuation Strength relied on Loudness and Sharpness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.015. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (15) has been determined. **Figure 132** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 133** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Fluctuation\ Strength = -6.594 - 1.991 \times Roughness + 0.096 \times LZeq \quad (14)$$

$$SDM\ Fluctuation\ Strength = -0.388 + 0.081 \times Loudness - 0.012 \times LAeq \quad (15)$$



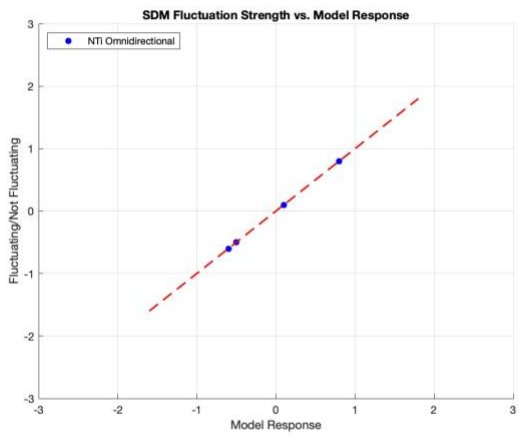


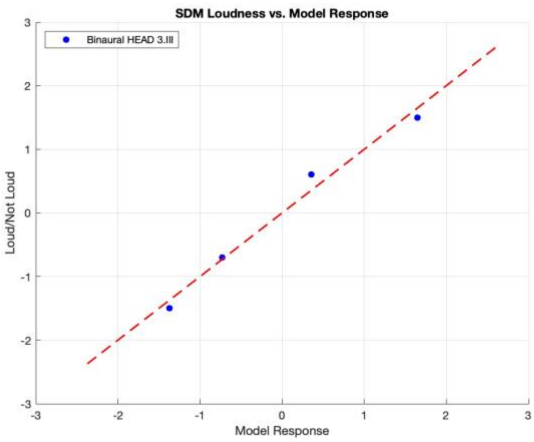
Fig 132-133 Comparison of subjective responses for the SDM Fluctuation Strength and the prediction model responses by Eq. (14) Eq. (15). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Loudness**, relative to **Table 19**, revealed a hierarchy of correlation strength, with LZeq demonstrating the highest correlation at 0.99, followed by LAeq at 0.98. After some tests, the best prediction model for noise loudness relied on LZeq, resulting in an **adjusted R-square** value of 0.97 and a Root Mean Square Error (**RMSE**) of 0.22. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (16) has been determined.

The analysis of the correlation reported in **table 20**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by Fluctuation Strength at 0.98. After some tests, the best prediction model for noise loudness relied on LAeq and Fluctuation Strength, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.1. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (17) has been determined. **Figure 134** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 135** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Loudness = -12.953 + 0.181 \times LZeq \tag{16}$$

$$SDM\ Loudness = -2.41 + 0.06 \times Loudness - 0.035 \times LAeq \tag{17}$$



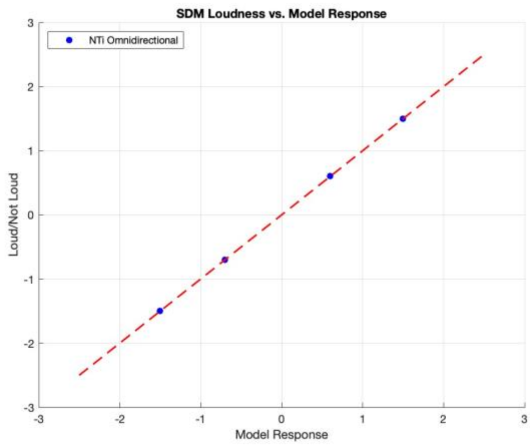


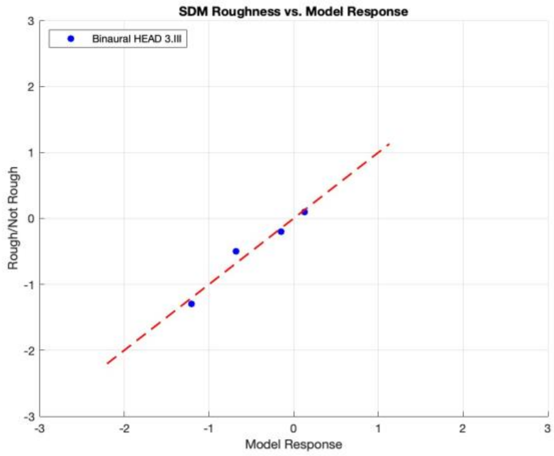
Fig 134-135 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (16) and Eq. (17). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 19**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.97, followed by LZeq at 0.91. After some tests, the best prediction model for noise roughness relied on LAeq, resulting in an **adjusted R-square** value of 0.94 and a Root Mean Square Error (**RMSE**) of 0.15. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (18) has been determined.

The analysis of the correlation reported in **table 20**, involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.98, followed by LAeq at 0.94. After some tests, the best prediction model for noise roughness relied on Roughness and Tonality, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.06. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (19) has been determined. **Figure 136** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 137** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Roughness = -2.123 + 0.033 \times LAeq \tag{18}$$

$$SDM\ Roughness = -2.17 + 1.14 \times Tonality + 11.73 \times Roughness \tag{19}$$



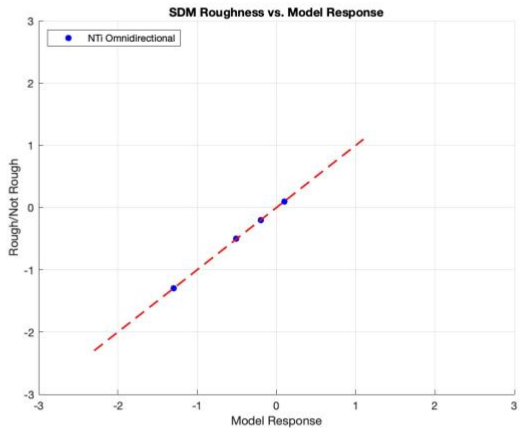


Fig 136-137 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (18) and Eq. (19). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.2 Immersive Test

Table 25 and **26** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	1	-3	-3	-3	-2.9	-3
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	1.9	-2.1	-1.7	-2.8	-1	-3
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	2.3	-0.7	-0.6	-1.4	-0.7	-2.4
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	2.9	-0.1	0.5	-1.7	-0.6	-2.1

Table 25 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	1	-3	-3	-3	-2.9	-3
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	1.9	-2.1	-1.7	-2.8	-1	-3
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	2.3	-0.7	-0.6	-1.4	-0.7	-2.4
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	2.9	-0.1	0.5	-1.7	-0.6	-2.1

Table 26 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.2.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 27**. In **Table 28** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.99307259	0.927301255	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.55091071	0.718642653	0.45897171	0.471573312	1								
Tonality	0.94444507	0.862469629	0.95647201	0.886317821	0.507296149	1							
Fluctuation Strength	0.9262928	0.969661811	0.88882791	0.953779286	0.610116675	0.755660283	1						
Annoyance Scale	0.95940216	0.985318111	0.92431508	0.918628154	0.763078923	0.909574253	0.913751164	1					
SDM Loudness	0.97317615	0.995972636	0.94192429	0.972578561	0.657041618	0.858181098	0.982631513	0.97010681	1				
SDM Roughness	0.98295608	0.993691846	0.95525738	0.958979459	0.691991127	0.913746037	0.944902321	0.992859456	0.98912921	1			
SDM Sharpness	0.8391176	0.904062118	0.79591671	0.895427253	0.539380933	0.615076438	0.97985842	0.81819318	0.92579346	0.86072605	1		
SDM Fluctuation Strength	0.77977103	0.901775397	0.70495016	0.725892213	0.948511481	0.702792896	0.822745865	0.921108228	0.86198391	0.88121077	0.75089603	1	
SDM Tonality	0.97023779	0.929442113	0.96917706	0.996701197	0.411498202	0.855559937	0.949744978	0.884829452	0.95719262	0.93402546	0.90351289	0.67893474	1

Table 27 Correlation coefficients for the immersive test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.96061386	1										
Sharpness	0.94089113	0.95585422	0.990153436	1									
Roughness	0.9398703	0.95816669	0.846888462	0.832203062	1								
Tonality	0.98274366	0.9379216	0.957775696	0.910490685	0.87514514	1							
Fluctuation Strength	0.94064849	0.98089541	0.953549564	0.974546922	0.90837314	0.8749913	1						
Annoyance Scale	0.99332588	0.99175766	0.943415233	0.922139452	0.9725917	0.95802937	0.94792055	1					
SDM Loudness	0.96742242	0.9929464	0.970497862	0.978848229	0.9243922	0.9148939	0.99587225	0.97010681	1				
SDM Roughness	0.993946	0.9973913	0.97378092	0.961702923	0.94565212	0.95962683	0.97205681	0.99285946	0.98912921	1			
SDM Sharpness	0.80002442	0.88365213	0.851144163	0.911825245	0.79347648	0.69931563	0.95615536	0.81819318	0.92579346	0.86072605	1		
SDM Fluctuation Strength	0.87099855	0.90302324	0.750918598	0.741637027	0.98638258	0.78735452	0.85101595	0.92110823	0.86198391	0.88121077	0.75089603	1	
SDM Tonality	0.91116262	0.92554169	0.982091588	0.995942736	0.77894922	0.88686174	0.95514557	0.88482945	0.95719262	0.93402546	0.90351289	0.67893474	1

Table 28 Correlation coefficients for the immersive test. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 27**, the analysis revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.98, followed by LZeq at 0.95. LZeq is not used due to its collinearity with LAeq. After some tests, the best prediction model for noise annoyance relied on LAeq and Fluctuation Strength, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.01. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (20) has been determined. **Figure 138** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -0.641 + 0.075 \times LAeq - 152.427 \times Fluctuation\,Strength \quad (20)$$

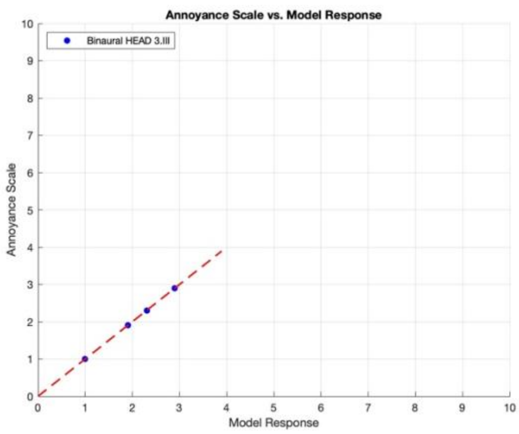


Fig 138 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (20). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 28**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.968, followed by Fluctuation Strength at 0.966. After some tests, the best prediction model for noise annoyance relied on LAeq, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.01. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (21) has been determined. **Figure 139** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -0.24 + 0.067 \times Laeq - 141.26 \times Fluctuation\ Strength \tag{21}$$

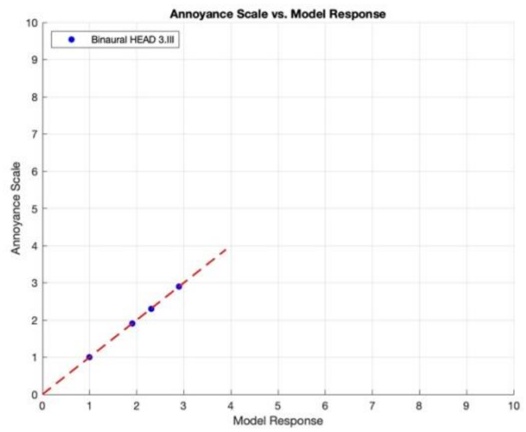


Fig 139 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (21). The straight red line indicates the cases where model responses are equal to subjective responses.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 29 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 30** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	1
V4	33.5	32.5	35	32.5	1.9
V8	50	50	52	47.5	2.3
V12	59	60	62	57	2.9

Table 29 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99977304	1			
PNC	0.99815903	0.999209045	1		
RNC	0.99660463	0.994984048	0.99037924	1	
Annoyance Scale	0.98315014	0.986197149	0.98960649	0.974723875	1

Table 30 Correlation coefficients for the immersive test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Preferred Noise Criterion, resulting in an **adjusted R-square** value of 0.97 and a Root Mean Square Error (**RMSE**) of 0.14. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (22) has been determined. **Figure 140** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are worse than the ones reported above, showing that the parameters based on spectral shape are not a better indicator of the annoyance prediction model.

$$Annoyance = 0.434 + 0.0387 \times PNC \quad (22)$$

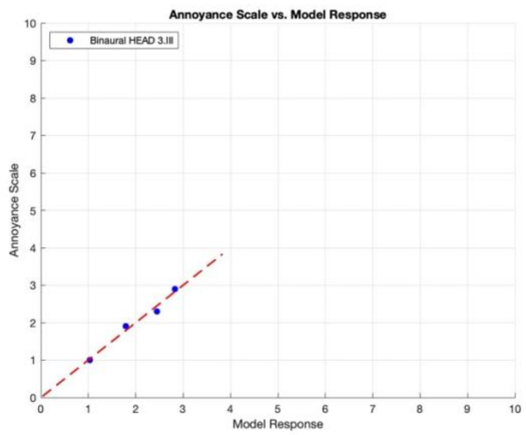


Fig 140 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (22). The straight red line indicates the cases where model responses are equal to subjective responses

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 31** shows the data used for the prediction and **table 32** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	1
V4	25	24	23	25	1.9
V8	42	41	42	40	2.3
V12	50	52	51	48	2.9

Table 31 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.998367538	1		
RNC	0.99975084	0.999383483	0.99975183	1	
Annoyance Scale	0.98291023	0.986569943	0.98314983	0.984436053	1

Table 32 Correlation coefficients for the immersive test. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Preferred Noise Criterion, resulting in an **adjusted R-square** value of 0.96 and a Root Mean Square Error (**RMSE**) of 0.16. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (23) has been determined. **Figure 141** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are better than the ones reported above, showing that the parameters based on spectral shape are a better indicator of the annoyance prediction model.

$$Annoyance = 0.752 + 0.041 \times PNC \tag{23}$$

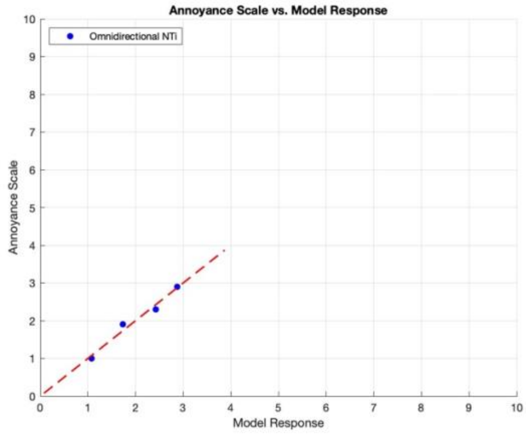


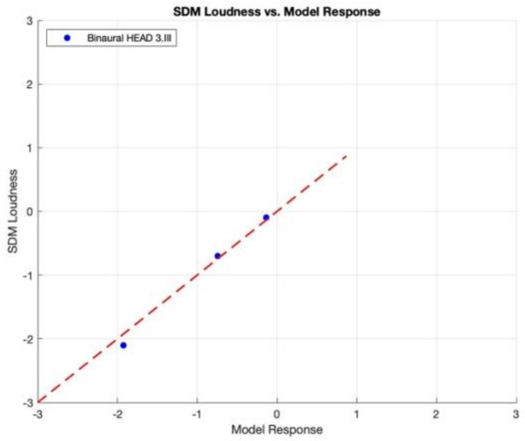
Fig 141 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (23). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.2.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 27** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 28** the ones from the omnidirectional microphone NTi.
Starting with the data reported in **table 27**, the analysis for the subjective parameter **Loudness** revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by Fluctuation Strength at 0.98. After some tests, the best prediction model for noise loudness relied on LAeq, resulting in an **adjusted R-square** value of 0.987 and a Root Mean Square Error (**RMSE**) of 0.14. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (24) has been determined.
The analysis of the correlation reported in **table 28**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by Fluctuation Strength at 0.97. After some tests, the best prediction model for noise loudness relied on LAeq, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.13. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (25) has been determined. **Figure 142** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 143** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Loudness = -5.156 + 0.0740 \times LAeq \quad (24)$$

$$SDM\ Loudness = 0.08 + 0.04 \times LAeq \quad (25)$$



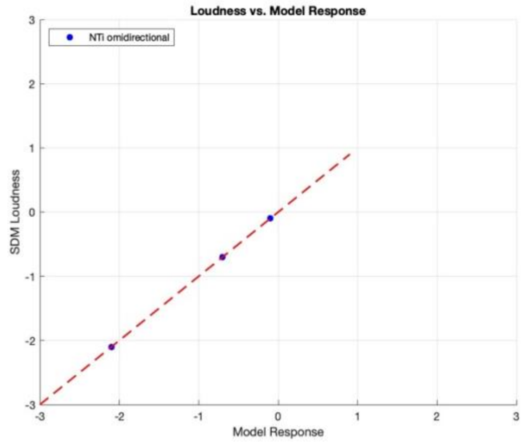
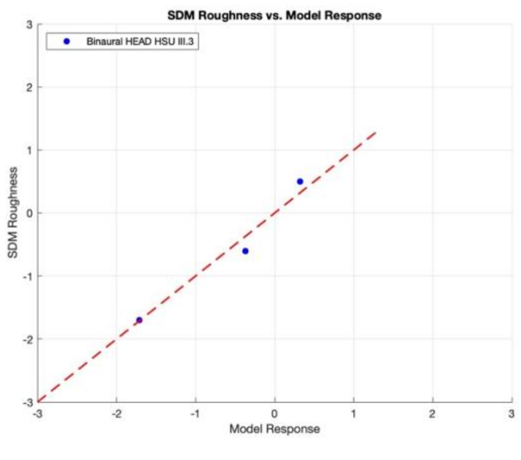


Fig 142-143 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (24) Eq. (25). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 27**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by LZeq at 0.98. After some tests, the best prediction model for noise roughness relied on LAeq, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.21. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (26) has been determined.
The analysis of the correlation reported in **table 28**, involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by Fluctuation Strength at 0.98. After some tests, the best prediction model for noise roughness relied on LAeq, resulting in an **adjusted R-square** value of 0.97 and a Root Mean Square Error (**RMSE**) of 0.26. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (27) has been determined. **Figure 144** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 145** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Roughness = \ -5.379\ +\ 0.084\ \times\ Laeq \quad (26)$$

$$SDM\ Roughness = -5.186\ +\ 0.098\ \times\ Laeq \quad (27)$$



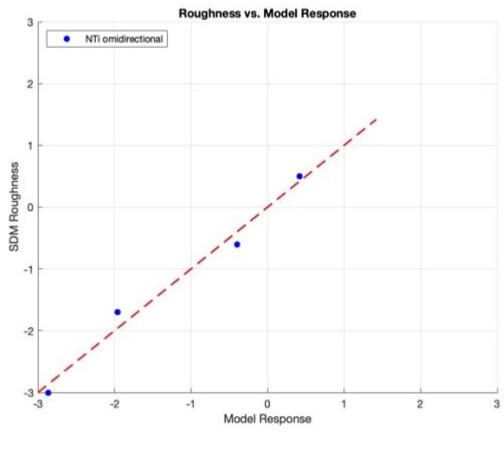


Fig 144-145 Comparison of subjective responses for the SDM Roughness and the prediction model responses by Eq. (26) and Eq. (27). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Sharpness**, relative to **Table 27**, revealed a hierarchy of correlation strength, with fluctuation demonstrating the highest correlation at 0.98, followed by LAeq at 0.90. After some tests, the best prediction model for sharpness relied on fluctuation strength, resulting in an **adjusted R-square** value of 0.94 and a Root Mean Square Error (**RMSE**) of 0.19. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (28) has been determined. . **Figure 146** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.
The analysis for the subjective parameter **Sharpness**, relative to **Table 27**, revealed a hierarchy of correlation strength too low to be predicted.

$$SDM\ Sharpness = -3.694 + 213.035 \times Fluctuation\ Strength \quad (28)$$

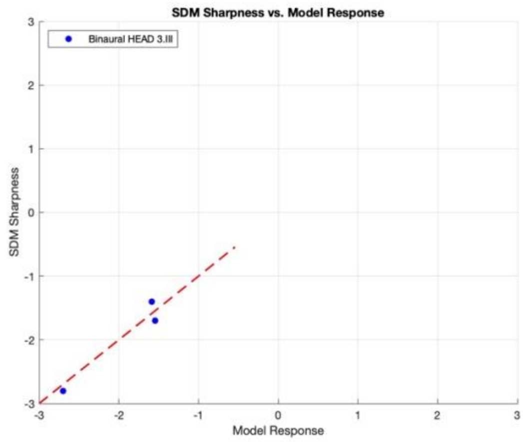


Fig 146 Comparison of subjective responses for the SDM Sharpness and the prediction model responses by Eq. (28). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Fluctuation Strength**, relative to **Table 27**, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.95, followed by LAeq at 0.90. After some tests, the best prediction model for fluctuation strength relied on LAeq and Roughness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.019. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (29) has been determined.
The analysis of the correlation reported in **table 28**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.99, followed by LAeq at 0.84. After some tests, the best prediction model for fluctuation strength relied on Roughness, resulting in an **adjusted R-square** value of 0.987 and a Root Mean Square Error (**RMSE**) of 0.12. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (30) has been determined. **Figure 147** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 148** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Fluctuation\ Strength = -4.465 + 12.184 \times Roughness + 0.028 \times Laeq \quad (29)$$

$$SDM\ Fluctuation\ Strength = -4.902 + 30.995 \times Roughness \quad (30)$$

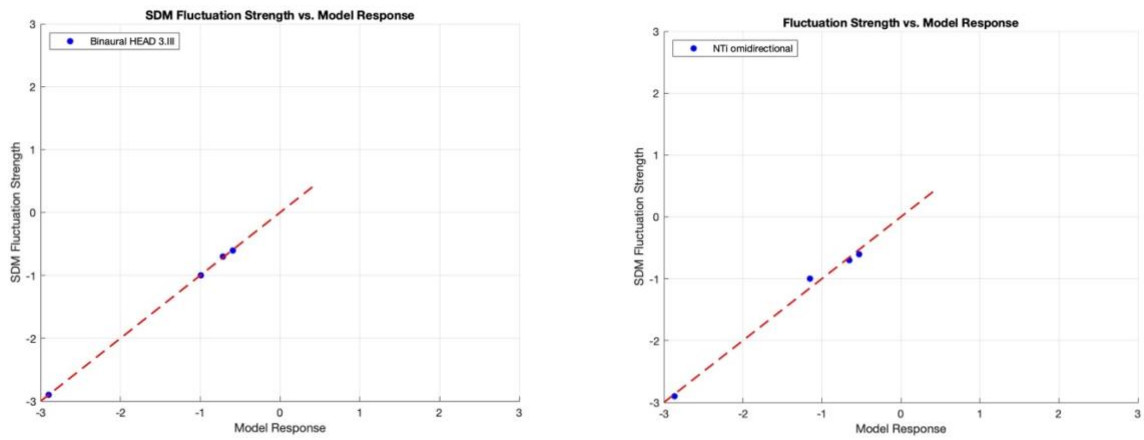


Fig 147-148 Comparison of subjective responses for the SDM fluctuation strength and the prediction model responses by Eq. (29) and Eq. (30). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Tonality**, relative to **Table 27**, revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.99, followed by LZeq at 0.97. After some tests, the best prediction model for noise tonality relied on Sharpness, resulting in an **adjusted R-square** value of 0.990 and a Root Mean Square Error (**RMSE**) of 0.04. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (31) has been determined.
The analysis of the correlation reported in **table 28**, involving **Tonality** and the NTi values, revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.99, followed by LZeq at 0.99. After some tests, the best prediction model for noise tonality relied on Sharpness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.05. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (32) has been determined. **Figure 149** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 150** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Tonality = -3.498 + 0.743 \times Sharpness \quad (31)$$

$$SDM\ Tonality = 3.42 + 0.71 \times Sharpness \quad (32)$$

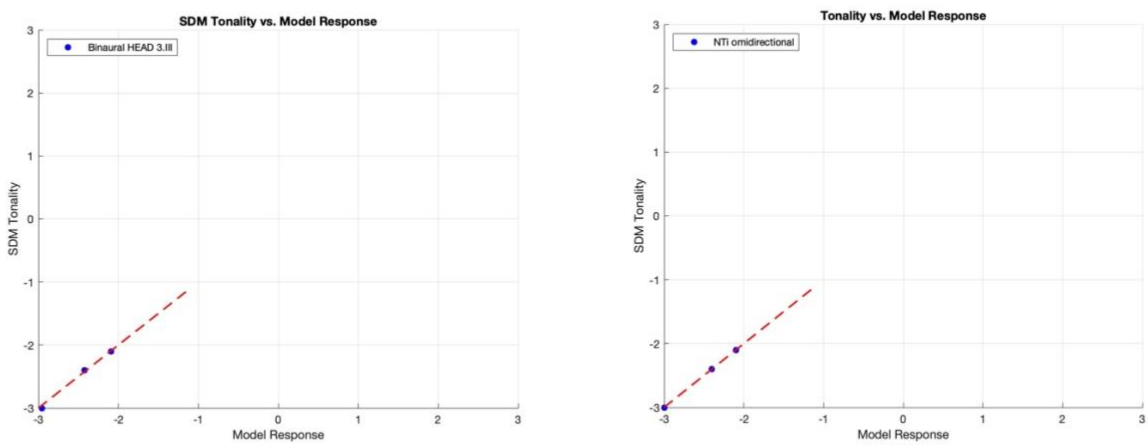


Fig 149-150 Comparison of subjective responses for the SDM Tonality and the prediction model responses by Eq. (31) and Eq. (32). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.3 Immersive Listening Test with Headphones

Table 33 and **34** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	3	-1.6	0.2	0	-1.2	1
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	3.8	-1.4	1	-1	-1.2	-1.4
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	5.2	0.2	2	-0.8	-1.8	-2
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	3.8	-0.6	0.8	0	-1.4	-0.8

Table 33 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	3	-1.6	0.2	0	-1.2	1
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	3.8	-1.4	1	-1	-1.2	-1.4
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	5.2	0.2	2	-0.8	-1.8	-2
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	3.8	-0.6	0.8	0	-1.4	-0.8

Table 34 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.3.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 35**. In **Table 36** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.993072594	0.927301255	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.550910711	0.718642653	0.45897171	0.471573312	1								
Tonality	0.94444507	0.862469629	0.95647201	0.886317821	0.507296149	1							
Fluctuation Strength	0.926292804	0.969661811	0.88882791	0.953779286	0.610116675	0.755660283	1						
Annoyance Scale	0.443984619	0.643713379	0.3493051	0.498826242	0.667534877	0.168821386	0.735912912	1					
SDM Loudness	0.678935209	0.79242218	0.61899433	0.755385598	0.520105027	0.404401323	0.904606175	0.917034059	1				
SDM Roughness	0.378791335	0.596551266	0.2767458	0.422517891	0.705958642	0.115974783	0.675170766	0.993433191	0.866397764	1			
SDM Sharpness	**0.191642104**	**-0.063414274**	**0.30373252**	**0.232952114**	**-0.683387538**	**0.282320262**	**-0.029789843**	-0.588901986	-0.223414282	-0.677479721	1		
SDM Fluctuation Strength	-0.501578609	-0.634495681	-0.4385094	-0.602474544	-0.400088107	-0.191589981	-0.786891407	-0.92766253	-0.974216034	-0.881917104	0.268866429	1	
SDM Tonality	**-0.478153496**	**-0.68996988**	**-0.3717649**	**-0.451158447**	**-0.937026256**	**-0.325328894**	**-0.663170471**	-0.87720776	-0.71280635	-0.907264709	0.762729044	0.654653671	1

Table 35 Correlation coefficients for the immersive listening test with headphones. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.96061386	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.95816669	0.84688846	0.83220306	1								
Tonality	0.98274366	0.9379216	0.9577757	0.91049069	0.87514514	1							
Fluctuation Strength	0.94064849	0.98089541	0.95354956	0.97454692	0.90837314	0.8749913	1						
Annoyance Scale	0.47162858	0.61209263	0.44351334	0.53326976	0.65911715	0.30085676	0.69264242	1					
SDM Loudness	0.64458106	0.76299075	0.69317319	0.7798548	0.70022774	0.50972694	0.86252776	0.91703406	1				
SDM Roughness	0.42592032	0.56608895	0.37311168	0.45847708	0.64501897	0.25133862	0.63492407	0.99343319	0.86639776	1			
SDM Sharpness	0.03811332	-0.0508802	0.2280324	0.20010364	-0.3051977	0.1746103	-0.0221048	-0.588902	-0.2234143	-0.6774797	1		
SDM Fluctuation Strength	-0.4554809	-0.5976184	-0.5220655	-0.6311114	-0.5371684	-0.3037856	-0.7269197	-0.9276625	-0.974216	-0.8819171	0.26886643	1	
SDM Tonality	-0.5880773	-0.677046	-0.4500277	-0.4810025	-0.8234326	-0.4462928	-0.662908	-0.8772078	-0.7128063	-0.9072647	0.76272904	0.65465367	1

Table 36 Correlation coefficients for the immersive listening test with headphones. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 35**, the analysis revealed a hierarchy of correlation strength, with Fluctuation Strength demonstrating the highest correlation at 0.74, followed by Roughness at 0.67. After some tests, the best prediction model for noise annoyance relied on Sharpness and Fluctuation Strength, resulting in an **adjusted R-square** value of 0.995 and a Root Mean Square Error (**RMSE**) of 0.07. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (33) has been determined. **Figure 151** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = 2.974 + 722.237 \times Fluctuation\ Strength - 3.407 \times Sharpness \quad (33)$$

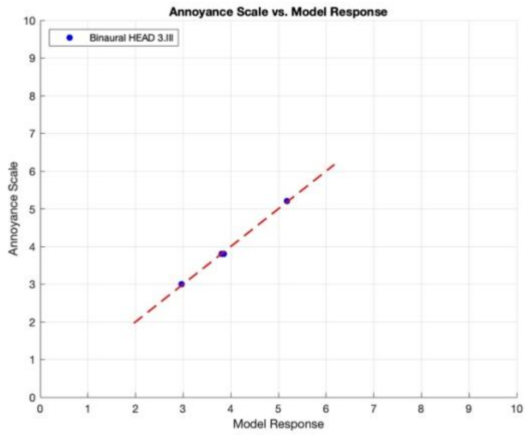


Fig 151 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (33). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 36**, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.72, followed by LAeq at 0.6. After some tests, the best prediction model for noise annoyance relied on LZeq and Tonality, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.1. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (34) has been determined. **Figure 152** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = 20.55 + 0.48 \times LZeq - 44.12 \times Tonality \quad (34)$$

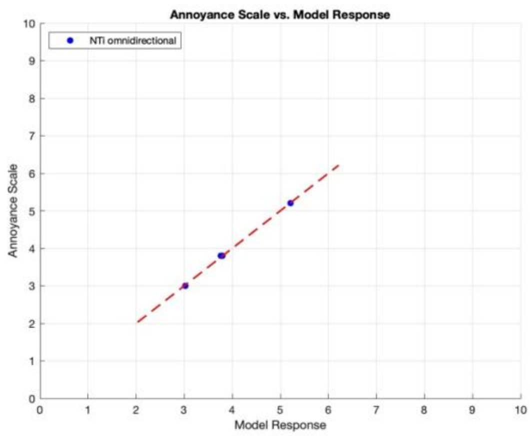


Fig 152 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (34). The straight red line indicates the cases where model responses are equal to subjective responses.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room

Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 37 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 38** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	3
V4	33.5	32.5	35	32.5	3.8
V8	50	50	52	47.5	5.2
V12	59	60	62	57	3.8

Table 37 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.999773043	1			
PNC	0.998159027	0.999209045	1		
RNC	0.996604626	0.994984048	0.99037924	1	
Annoyance Scale	0.628225418	0.629474096	0.63825682	0.583090416	1

Table 38 Correlation coefficients for the immersive listening test with headphones. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the spectrum-based parameters and the annoyance is too low to be used to calculate the linear regression.

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 39** shows the data used for the prediction and **table 40** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	3
V4	25	24	23	25	3.8
V8	42	41	42	40	5.2
V12	50	52	51	48	3.8

Table 39 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.998367538	1		
RNC	0.99975084	0.999383483	0.99975183	1	
Annoyance Scale	0.65236981	0.608592133	0.6520297	0.635860213	1

Table 40 Correlation coefficients for the immersive listening test with headphones. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the spectrum-based parameters and the annoyance is too low to be used to calculate the linear regression.

3.3.3.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 35** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 35** the ones from the omnidirectional microphone NTi.
Starting with the data reported in **table 35**, the analysis for the subjective parameter **Loudness** revealed a hierarchy of correlation strength, with Fluctuation strength demonstrating the highest correlation at 0.90, followed by LAeq at 0.79. After some tests, the best prediction model for noise loudness relied on Fluctuation strength and tonality, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**)

of 0.003. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (35) has been determined. **Figure 153** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.
The analysis of the correlation reported in **table 36**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength too low to have a reliable predictive model.

$$SDM\ Loudness = -2.160 - 4.208 \times Tonality + 314.843 \times Fluctuation\ Strength \quad (35)$$

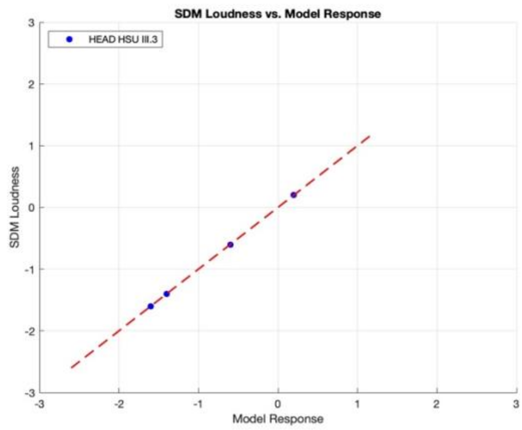


Fig 153 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (35). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 35**, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.71. After some tests, the best prediction model for noise roughness relied on Sharpness and Fluctuation strength, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.04. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (36) has been determined.
The analysis of the correlation reported in **table 36**, involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.72. After some tests, the best prediction model for noise roughness relied on LZeq and Tonality, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.05. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (37) has been determined. **Figure 154** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 155** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Roughness = -14.667 + 0.295 \times Lzeq - 23.043 \times Tonality \quad (36)$$

$$SDM\ Roughness = -19.3 + 0.396 \times Lzeq - 37.12 \times Tonality \quad (37)$$

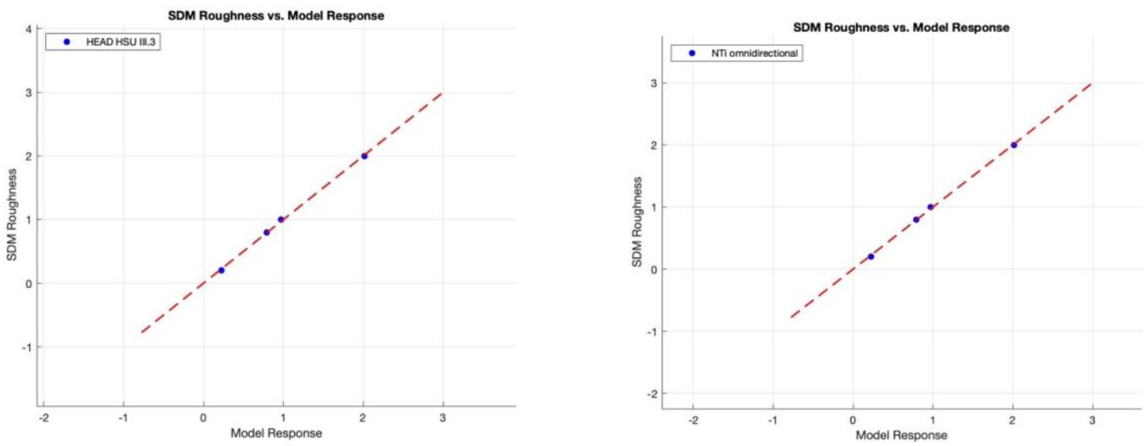


Fig 154-155 Comparison of subjective responses for the SDM Roughness and the prediction model responses by Eq. (36) and Eq. (37). The straight red line indicates the cases where model responses are equal to subjective responses.

The best prediction model for noise **sharpness** relied on Roughness and Tonality, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.014. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (36) has been determined.
The analysis of the correlation reported in **table 36**, involving **Sharpness** and the NTi values, revealed a hierarchy of correlation strength, with Tonality demonstrating the highest correlation at 0.61. After some tests, the best prediction model for noise sharpness relied on Roughness and LZeq, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.04. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (37) has been determined. **Figure 156** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 157** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Sharpness = 0.390 + 3.502 \times Tonality - 10.629 \times Roughness \quad (38)$$

$$SDM\ Sharpness = -6.02 + -36.06 \times Roughness + 0.14 \times LZeq \quad (39)$$

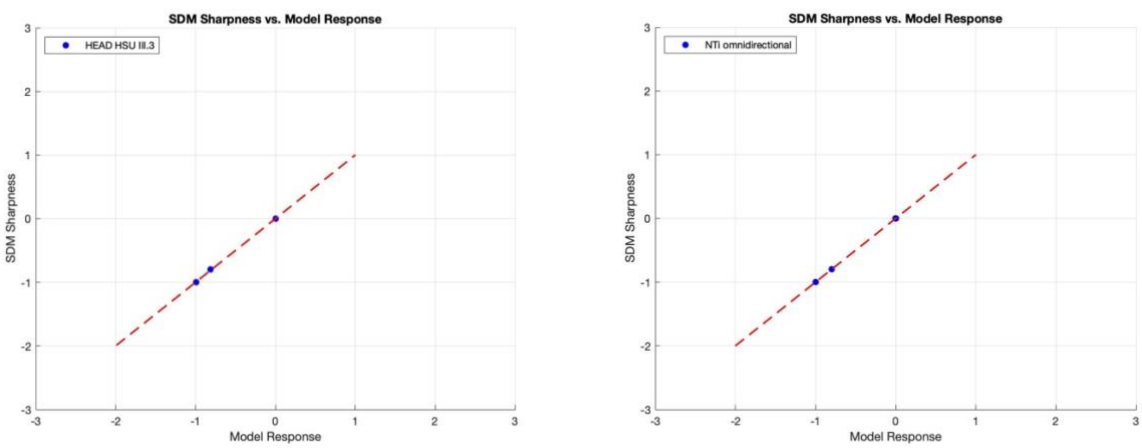


Fig 156-157 Comparison of subjective responses for the SDM Sharpness and the prediction model responses by Eq. (38) Eq. (39). The straight red line indicates the cases where model responses are equal to subjective responses.

After some tests, the best prediction model for **fluctuation strength** relied on fluctuation strength and tonality, resulting in an **adjusted R-square** value of 0.993 and a Root Mean Square Error (**RMSE**) of 0.02. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (40) has been determined. **Figure 158** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.
The analysis of the correlation reported in **table 36**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation not suitable for prediction purposes.

$$SDM\ Fluctuation\,Strength = -1.027 + -116.045 \times Fluctuation\,Strength + 2.089 \times Tonality \qquad (40)$$

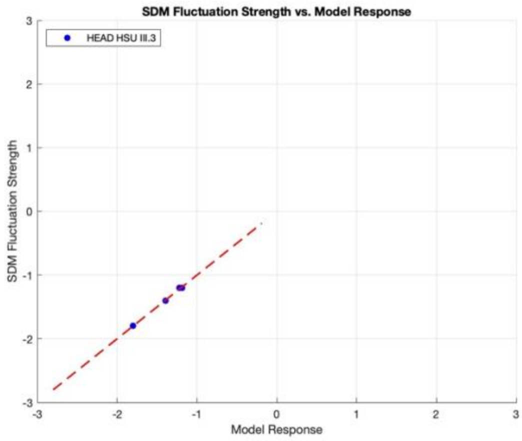
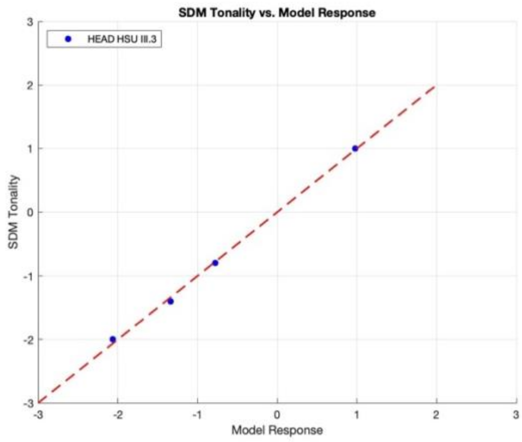


Fig 158 Comparison of subjective responses for the SDM fluctuation strength and the prediction model responses by Eq. (40). The straight red line indicates the cases where model responses are equal to subjective responses.

After some tests, the best prediction model for **tonality** relied on LZeq and Loudness, resulting in an **adjusted R-square** value of 0.994 and a Root Mean Square Error (**RMSE**) of 0.09. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (41) has been determined.
After some tests, the best prediction model for **tonality** relied on LZeq and Roughness, resulting in an **adjusted R-square** value of 0.92 and a Root Mean Square Error (**RMSE**) of 0.36. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (32) has been determined. **Figure 159** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 160** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Tonality = 89.602 + 0.793 \times Loudness - 1.403 \times Lzeq \qquad (41)$$

$$SDM\ Tonality = -6.26 + -70.53 \times Roughness + 0.21 \times LZeq \qquad (42)$$



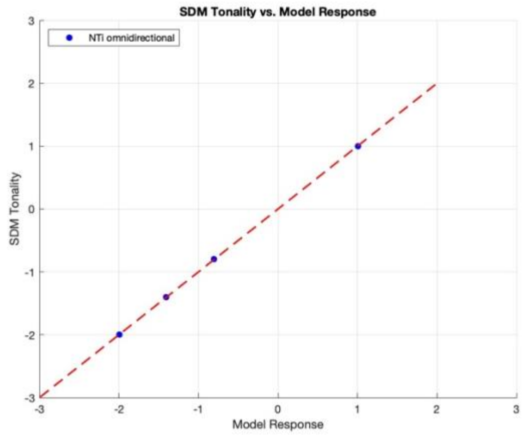


Fig 159-160 Comparison of subjective responses for the SDM Tonality and the prediction model responses by Eq. (41) and Eq. (42). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.4 Binaural Test

Table 41 and **42** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	1.6	-1.5	-1.7	-2.2	-2	-2.5
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	2.5	-1.5	-1.5	-1.5	-1.7	-1.7
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	4.7	0.3	-0.8	-1.7	-0.8	-1.8
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	6.2	0.8	0.5	-1.2	-0.8	-1.3

Table 41 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	1.6	-1.5	-1.7	-2.2	-2	-2.5
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	2.5	-1.5	-1.5	-1.5	-1.7	-1.7
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	4.7	0.3	-0.8	-1.7	-0.8	-1.8
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	6.2	0.8	0.5	-1.2	-0.8	-1.3

Table 42 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.4.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 43**. In **Table 44** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.993072594	0.927301255	1										
Sharpness	0.9862058	0.95164994	0.981158581	1									
Roughness	0.550910711	0.718642653	0.458971714	0.471573312	1								
Tonality	0.94444507	0.862469629	0.956472011	0.886317821	0.507296149	1							
Fluctuation Strength	0.926292804	0.969661811	0.888827912	0.953779286	0.610116675	0.755660283	1						
Annoyance Scale	0.993288625	0.976815998	0.979209114	0.994826666	0.557239666	0.900152888	0.962617812	1					
SDM Loudness	0.950340945	0.931106631	0.940888521	0.987998245	0.428272357	0.804277483	0.970088641	0.975895777	1				
SDM Roughness	0.982609489	0.899220129	0.9973432	0.96697205	0.414658099	0.965501981	0.853811747	0.961886878	0.920176001	1			
SDM Sharpness	0.820163237	0.840670752	0.786405027	0.719537126	0.821367227	0.886646739	0.683202718	0.778298338	0.623986704	0.77744794	1		
SDM Fluctuation Strength	0.917232822	0.965833899	0.877774288	0.946719439	0.613494749	0.740594598	0.999721456	0.95596511	0.965935827	0.841317452	0.673201417	1	
SDM Tonality	0.862220525	0.895806837	0.822785856	0.776955936	0.841950013	0.890035378	0.761385797	0.832783872	0.696391718	0.807391407	0.992970815	0.75311668	1

Table 43 Correlation coefficients for the binaural test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.96061386	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.95816669	0.84688846	0.83220306	1								
Tonality	0.98274366	0.9379216	0.9577757	0.91049069	0.87514514	1							
Fluctuation Strength	0.94064849	0.98089541	0.95354956	0.97454692	0.90837314	0.8749913	1						
Annoyance Scale	0.96805907	0.97675341	0.99484679	0.99567687	0.87469646	0.93887983	0.9791518	1					
SDM Loudness	0.8903836	0.92259496	0.96388393	0.99116178	0.78107207	0.84785796	0.96693237	0.97589578	1				
SDM Roughness	0.94467554	0.90930022	0.98465443	0.95680758	0.77788572	0.96673176	0.88637487	0.96188688	0.920176	1			
SDM Sharpness	0.91079614	0.86224778	0.78531117	0.71712186	0.91967846	0.91226847	0.74773678	0.77829834	0.6239867	0.77744794	1		
SDM Fluctuation Strength	0.89514017	0.95328588	0.92200313	0.95889696	0.87460578	0.81430013	0.99329715	0.95596511	0.96593583	0.84131745	0.67320142	1	
SDM Tonality	0.94328831	0.91244809	0.83115003	0.77776366	0.95837753	0.92759331	0.81699996	0.83278387	0.69639172	0.80739141	0.99297081	0.75311668	1

Table 44 Correlation coefficients for the binaural test. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 43**, the analysis revealed a hierarchy of correlation strength, with LZeq demonstrating the highest correlation at 0.99, followed by Sharpness at 0.99. After some tests, the best prediction model for noise annoyance relied on LZeq and Tonality, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.02. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (43) has been determined. **Figure 161** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -22.121 + 0.379 \times LZeq - 5.769 \times Tonality \quad (43)$$

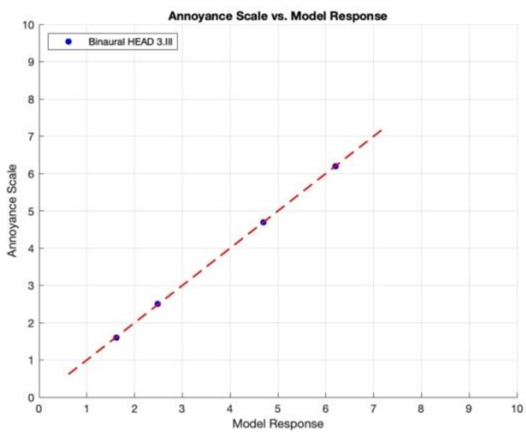


Fig 161 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (43). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 44**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by Fluctuation Strength at 0.98. After some tests, the best prediction model for noise annoyance relied on LAeq and Sharpness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.13. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (44) has been determined. **Figure 162** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -0.39 + 0.03 \times Laeq + 2.37 \times Sharpness \quad (44)$$

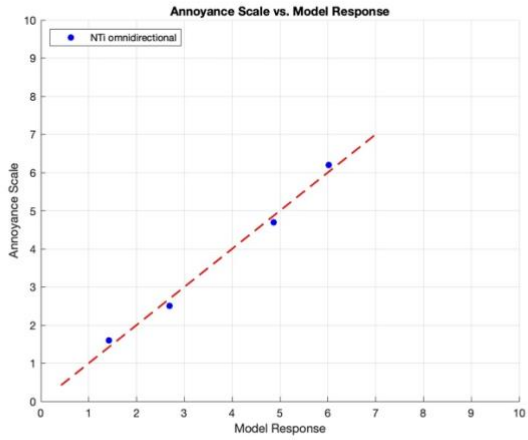


Fig 162 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (44). The straight red line indicates the cases where model responses are equal to subjective responses.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 45 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 46** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	1.6
V4	33.5	32.5	35	32.5	2.5
V8	50	50	52	47.5	4.7
V12	59	60	62	57	6.2

Table 45 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.999773043	1			
PNC	0.998159027	0.999209045	1		
RNC	0.996604626	0.994984048	0.990379244	1	
Annoyance Scale	0.983638597	0.98019485	0.971850064	0.995088197	1

Table 46 Correlation coefficients for the binaural test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Noise Criterion and Room Noise Criterion, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.003. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (45) has been determined. **Figure 163** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are better than the ones reported above, showing that the parameters based on spectral shape are a better indicator of the annoyance prediction model.

$$Annoyance = -2.826 - 0.105 \times NC + 0.269 \times RNC \quad (45)$$

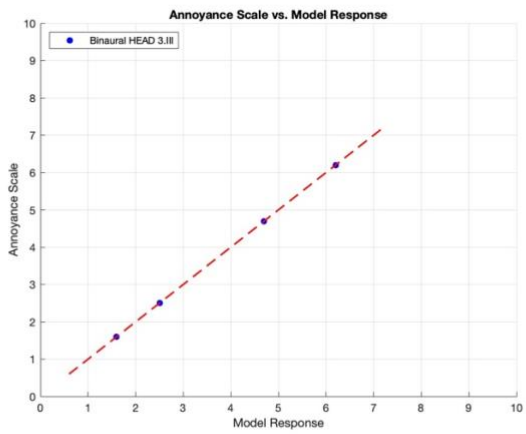


Fig 163 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (44). The straight red line indicates the cases where model responses are equal to subjective responses

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 47** shows the data used for the prediction and **table 48** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	1.6
V4	25	24	23	25	2.5
V8	42	41	42	40	4.7
V12	50	52	51	48	6.2

Table 47 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.998367538	1		
RNC	0.99975084	0.999383483	0.99975183	1	
Annoyance Scale	0.97668595	0.984595166	0.97652816	0.980374771	1

Table 48 Correlation coefficients for the binaural test. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Preferred Noise Criterion, resulting in an **adjusted R-square** value of 0.95 and a Root Mean Square Error (**RMSE**) of 0.45. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (45) has been determined. **Figure 164** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are worse than the ones reported above, showing that the parameters based on spectral shape are not a better indicator of the annoyance prediction model.

$$Annoyance = 0.422 + 0.106 \times PNC \quad (45)$$

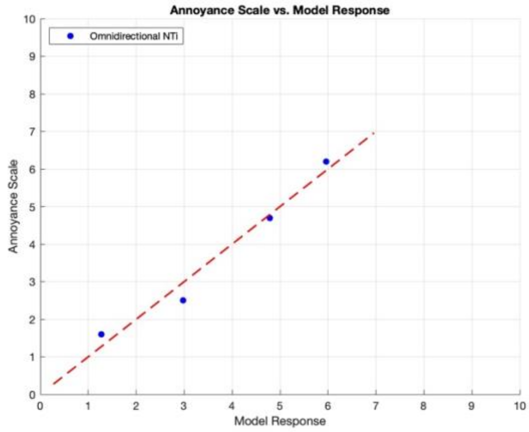


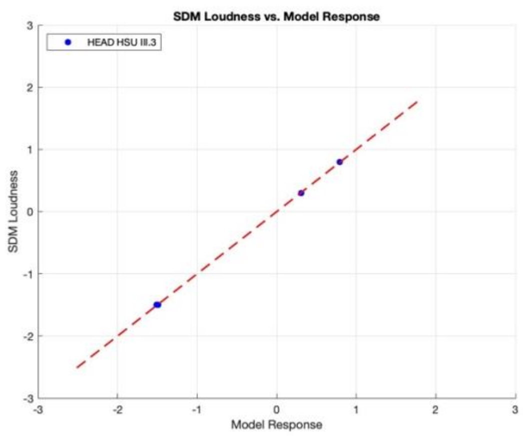
Fig 164 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (45). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.4.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 43** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 44** the ones from the omnidirectional microphone NTi.
Starting with the data reported in **table 43**, the analysis for the subjective parameter **Loudness** revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.98, followed by Fluctuation Strength at 0.97. After some tests, the best prediction model for noise loudness relied on Tonality and Sharpness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.02. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (46) has been determined.
The analysis of the correlation reported in **table 44**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.98, followed by LZeq at 0.97. After some tests, the best prediction model for noise loudness relied on Sharpness and Tonality, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.26. Importantly, all p-values were below 0.03, indicating the statistical significance of the findings. The multiple linear regression equation (47) has been determined. **Figure 165** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 166** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Loudness = -2.833 - 3.144 \times Tonality + 2.553 \times Sharpness \quad (46)$$

$$SDM\ Loudness = -2.49 + 2.43 \times Sharpness - 3.89 \times Tonality \quad (47)$$



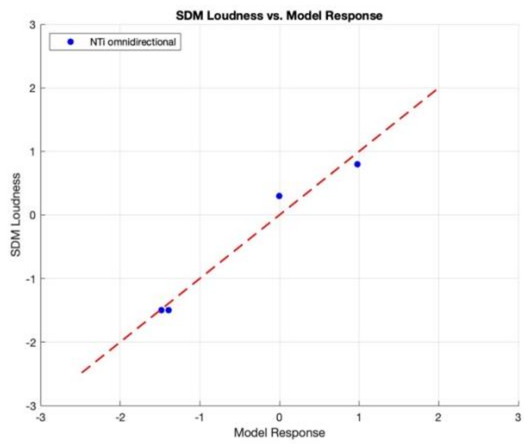
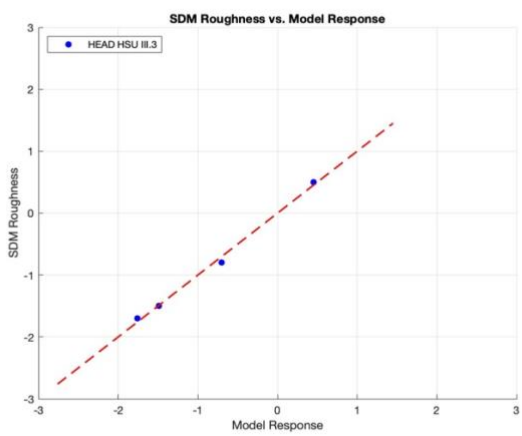


Fig 165-166 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (46) Eq. (47). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 43**, revealed a hierarchy of correlation strength, with Loudness demonstrating the highest correlation at 0.99, followed by LZeq at 0.98. After some tests, the best prediction model for noise roughness relied on Loudness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.088. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (48) has been determined.
The analysis of the correlation reported in **table 44**, involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with Loudness demonstrating the highest correlation at 0.99, followed by LZeq at 0.98. After some tests, the best prediction model for noise roughness relied on Loudness, resulting in an **adjusted R-square** value of 0.95 and a Root Mean Square Error (**RMSE**) of 0.21. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (49) has been determined. **Figure 167** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 168** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Roughness = -1.856 + 0.081 \times Loudness \quad (48)$$

$$SDM\ Roughness = -1.9 + 0.09 \times Loudness \quad (49)$$



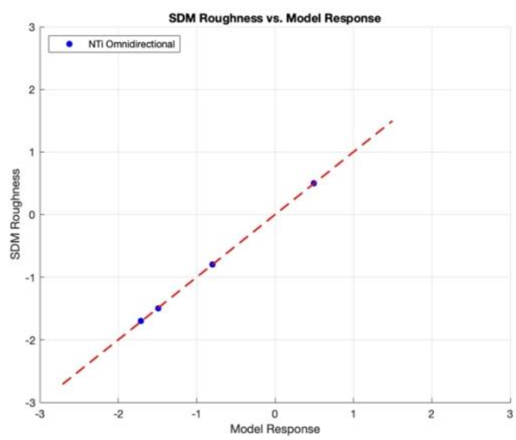


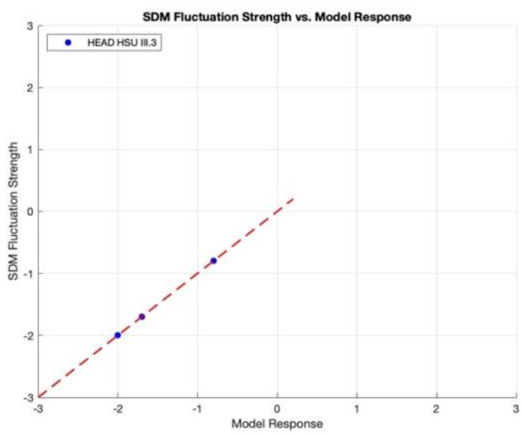
Fig 167-168 Comparison of subjective responses for the SDM Roughness and the prediction model responses by Eq. (48) and Eq. (49). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Sharpness**, relative to **Table 43**, revealed a hierarchy of correlation strength too low to be predicted.
The analysis for the subjective parameter **Sharpness**, relative to **Table 44**, revealed a hierarchy of correlation strength too low to be predicted.

The analysis for the subjective parameter **Fluctuation Strength**, relative to **Table 43**, revealed a hierarchy of correlation strength, with Fluctuation strength demonstrating the highest correlation at 0.99, followed by LAeq at 0.96. After some tests, the best prediction model for fluctuation strength relied on Fluctuation Strength and Loudness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.001. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (51) has been determined.
The analysis of the correlation reported in **table 44**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.97, followed by Sharpness at 0.92. After some tests, the best prediction model for fluctuation strength relied on Fluctuation Strength and Loudness, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.08. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (52) has been determined. **Figure 170** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 171** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Fluctuation\ Strength = -2.516 + 177.226 \times Fluctuation\ Strength - 0.003 \times Loudness \quad (51)$$

$$SDM\ Fluctuation\ Strength = -2.53 - 0.016 \times Loudness + 209.5 \times Fluctuation\ Strength \quad (52)$$



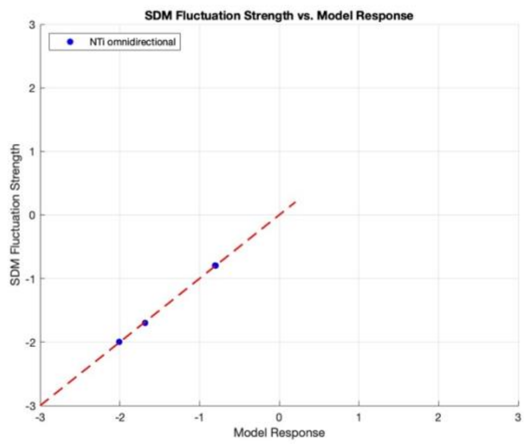


Fig 170-171 Comparison of subjective responses for the SDM fluctuation strength and the prediction model responses by Eq. (51) and Eq. (52). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.5 Compared Tests – Immersive Test

Table 49 and **50** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU

III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	1	-3	-3	-3	-3	-3
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	2.6	-1.4	-1.6	-2.8	-0.8	-3
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	2	-0.4	-0.8	-1.8	-0.8	-2.6
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	3.2	-0.4	0.4	-1.6	-0.6	-2

Table 49 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	1	-3	-3	-3	-3	-3
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	2.6	-1.4	-1.6	-2.8	-0.8	-3
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	2	-0.4	-0.8	-1.8	-0.8	-2.6
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	3.2	-0.4	0.4	-1.6	-0.6	-2

Table 50 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.5.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 51**. In **Table 52** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.99307259	0.92730126	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.55091071	0.71864265	0.45897171	0.471573312	1								
Tonality	0.94444507	0.86246963	0.95647201	0.886317821	0.507296149	1							
Fluctuation Strength	0.9262928	0.96966181	0.88882791	0.953779286	0.610116675	0.75566028	1						
Annoyance Scale	0.78685277	0.80155939	0.75580019	0.67751944	0.8082929	0.87510182	0.63132095	1					
SDM Loudness	0.86381942	0.96541594	0.79863896	0.84484059	0.852403908	0.73518491	0.93422665	0.78740419	1				
SDM Roughness	0.97696827	0.9857245	0.94946266	0.941042469	0.714494609	0.92936544	0.91937374	0.87643577	0.93240857	1			
SDM Sharpness	0.95284469	0.96394989	0.92922861	0.981457059	0.535894356	0.80017287	0.99251441	0.62731961	0.89699064	0.92447109	1		
SDM Fluctuation Strength	0.73075634	0.85614789	0.65434071	0.660054511	0.972170281	0.68405805	0.7535813	0.88736136	0.93605418	0.85849904	0.7010719	1	
SDM Tonality	0.97230562	0.87875211	0.99262498	0.968997518	0.348298143	0.94312501	0.85056804	0.69174888	0.72404228	0.90446774	0.90380398	0.55819925	1

Table 51 Correlation coefficients for the Compare test - immersive test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.960613856	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.958166689	0.84688846	0.832203062	1								
Tonality	0.98274366	0.937921602	0.9577757	0.910490685	0.875145144	1							
Fluctuation Strength	0.94064849	0.980895409	0.95354956	0.974546922	0.908373138	0.874991297	1						
Annoyance Scale	0.88376835	0.825801292	0.74971548	0.673479776	0.891792831	0.894311454	0.700592874	1					
SDM Loudness	0.91166268	0.959936277	0.84809754	0.860659862	0.9808083	0.821545904	0.947306863	0.787404193	1				
SDM Roughness	0.99852341	0.992500896	0.96409871	0.942521311	0.953450181	0.971264316	0.95302792	0.876435772	0.93240857	1			
SDM Sharpness	0.91464357	0.954917975	0.96070798	0.988281577	0.846118304	0.856809289	0.991049905	0.62731961	0.89699064	0.92447109	1		
SDM Fluctuation Strength	0.84022732	0.861224144	0.69616171	0.674808806	0.970647825	0.764165308	0.789716121	0.887361361	0.93605418	0.85849904	0.7010719	1	
SDM Tonality	0.91842542	0.886765906	0.97927391	0.958309432	0.734617031	0.94165654	0.877445487	0.691748877	0.72404228	0.90446774	0.90380398	0.55819925	1

Table 52 Correlation coefficients for the Compare test - immersive test. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 51**, the analysis revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.98, followed by LZeq at 0.95. LZeq is not used due to its collinearity with LAeq. After some tests, the best prediction model for noise annoyance relied on LAeq and Fluctuation Strength, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.05.

Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (53) has been determined. **Figure 172** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -1.810 + 0.168 \times LAeq - 627.885 \times Fluctuation\ Strength \quad (53)$$

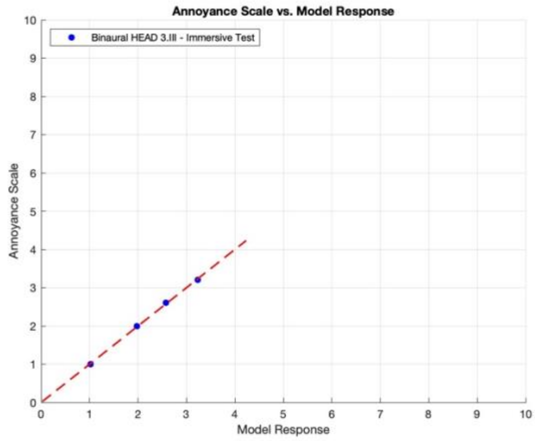


Fig 172 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (53). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 52**, revealed a hierarchy of correlation strength too low to be suitable for the prediction model.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 53 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 54** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	1
V4	33.5	32.5	35	32.5	2.6
V8	50	50	52	47.5	2
V12	59	60	62	57	3.2

Table 53 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99977304	1			
PNC	0.99815903	0.99920904	1		
RNC	0.99660463	0.99498405	0.99037924	1	
Annoyance Scale	0.79337611	0.8038889	0.8187588	0.775036367	1

Table 54 Correlation coefficients for the Compare test - immersive test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the subjective and objective parameters is too low to be suitable for the linear regression.

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 55** shows the data used for the prediction and **table 56** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	1
V4	25	24	23	25	2.6
V8	42	41	42	40	2
V12	50	52	51	48	3.2

Table 55 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.998367538	1		
RNC	0.99975084	0.999383483	0.99975183	1	
Annoyance Scale	0.79367219	0.805945404	0.79449976	0.797771164	1

Table 56 Correlation coefficients for the Compare test - immersive test. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the subjective and objective parameters is too low to be suitable for the linear regression.

3.3.5.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 51** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 52** the ones from the omnidirectional microphone NTi.
Starting with the data reported in **table 51**, the analysis for the subjective parameter **Loudness** revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.96, followed by Fluctuation Strength at 0.93. After some tests, the best prediction model for noise loudness relied on LAeq, resulting in an **adjusted R-square** value of 0.9 and a Root Mean Square Error (**RMSE**) of 0.39. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (53) has been determined.
The analysis of the correlation reported in **table 52**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.98, followed by LAeq at 0.93. After some tests, the best prediction model for noise loudness relied on Roughness, resulting in an **adjusted R-square** value of 0.96 and a Root Mean Square Error (**RMSE**) of 0.23. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (25) has been determined. **Figure 172** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 173** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Loudness = -4.624 + 0.067 \times Laeq \quad (53)$$

$$SDM\ Loudness = -5.362 + 34.961 \times Roughness \quad (54)$$

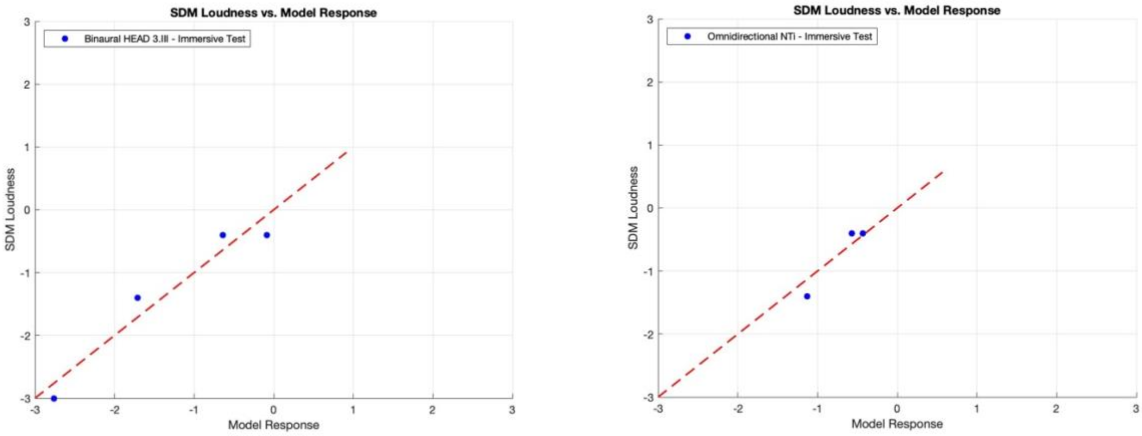


Fig 172-173 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (53) Eq. (54). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 51**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.98, followed by LZeq at 0.97. After some tests, the best prediction model for noise roughness relied on LAeq, resulting in an **adjusted R-square** value of 0.95 and a Root Mean Square Error (**RMSE**) of 0.29. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (55) has been determined.
The analysis of the correlation reported in **table 52** involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with Fluctuation Strength demonstrating the highest correlation at 0.98, followed by LAeq at 0.97. After some tests, the best prediction model for noise roughness relied on Fluctuation Strength, resulting in an **adjusted R-square** value of 0.95 and a Root Mean Square Error (**RMSE**) of 0.33. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (56) has been determined. **Figure 174** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 175** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Roughness = \ -5.197\ +\ 0.079\ \times\ Laeq \tag{55}$$

$$SDM\ Roughness = -4.246\ +\ 635.765\ \times Fluctuation\ Strength \tag{56}$$

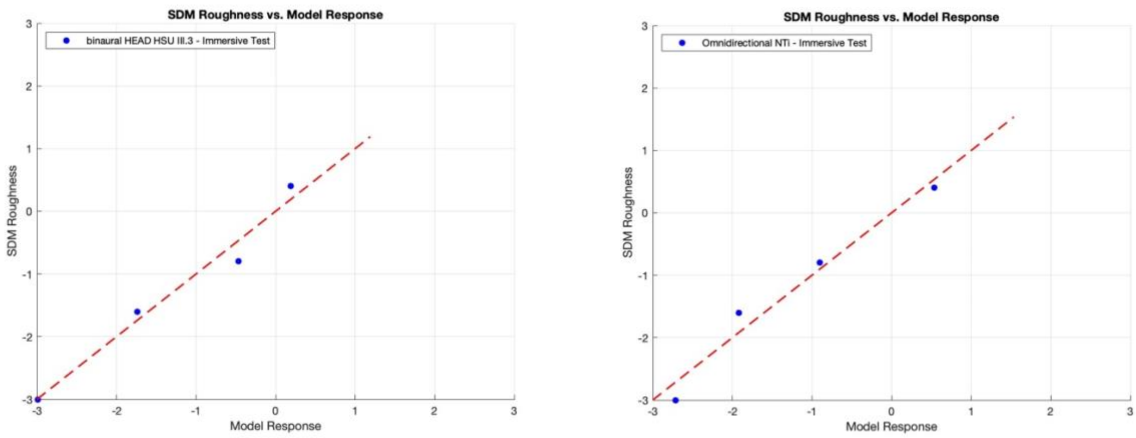


Fig 174-175 Comparison of subjective responses for the SDM Roughness and the prediction model responses by Eq. (55) and Eq. (56). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Sharpness**, relative to **Table 51**, revealed a hierarchy of correlation strength, with fluctuation strength demonstrating the highest correlation at 0.99, followed by Sharpness at 0.98. After some tests, the best prediction model for sharpness relied on fluctuation strength, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.1. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (57) has been determined. The analysis of the correlation reported in **table 52** involving **Sharpness** and the NTi values, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.98, followed by Sharpnessat 0.97. After some tests, the best prediction model for sharpness relied on Tonality and Sharpness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.0001. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (58) has been determined. **Figure 176** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 177** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Sharpness = -3.612 + 191.078 \times Fluctuation\ Strength \quad (57)$$

$$SDM\ Sharpness = -3.673 - 5.251 \times Tonality + 2.066 \times Sharpness \quad (58)$$

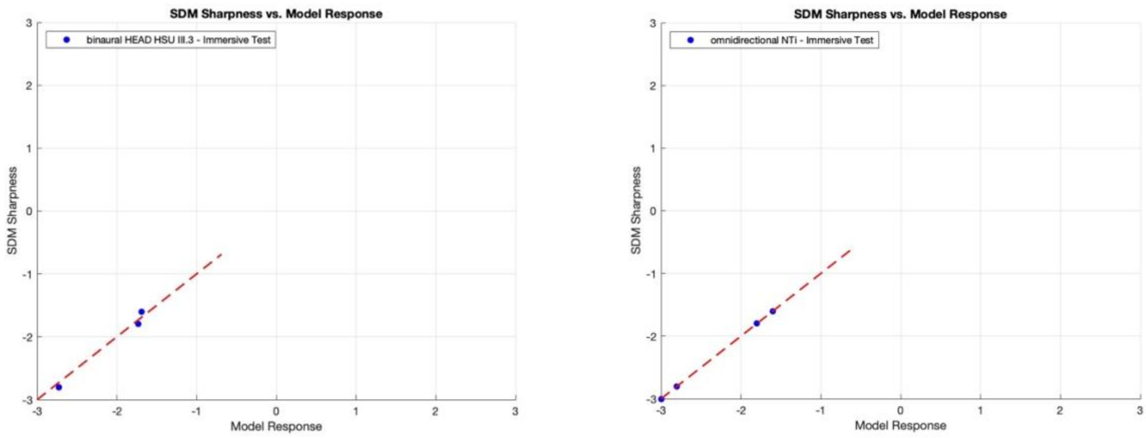


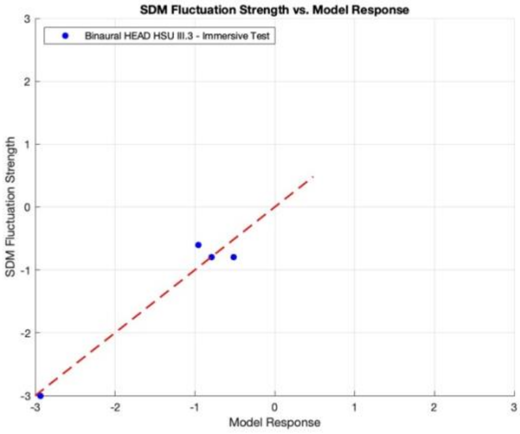
Fig 176-177 Comparison of subjective responses for the SDM Sharpness and the prediction model responses by Eq. (57) Eq. (58). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Fluctuation Strength**, relative to **Table 51**, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.97, followed by LAeq at 0.86. After some tests, the best prediction model for fluctuation strength relied on Roughness, resulting in an **adjusted R-square** value of 0.92 and a Root Mean Square Error (**RMSE**) of 0.33. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (59) has been determined.

The analysis of the correlation reported in **table 52**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.98, followed by LAeq at 0.79. After some tests, the best prediction model for fluctuation strength relied on Roughness, resulting in an **adjusted R-square** value of 0.94 and a Root Mean Square Error (**RMSE**) of 0.27. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (60) has been determined. **Figure 178** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 179** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Fluctuation\ Strength = \ -4.24\ +\ 20.071\ \times\ Roughness \tag{59}$$

$$SDM\ Fluctuation\ Strength = \ -5.028 +\ 32.08\ \times\ Roughness \tag{60}$$



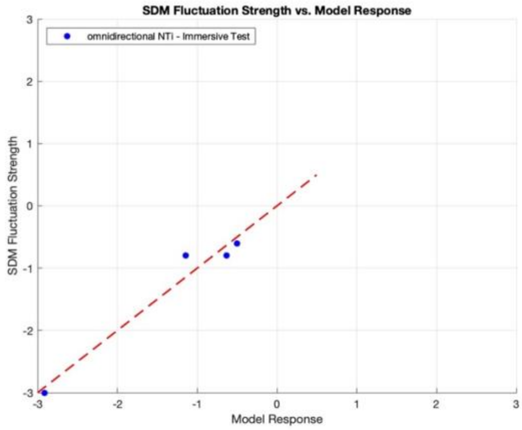
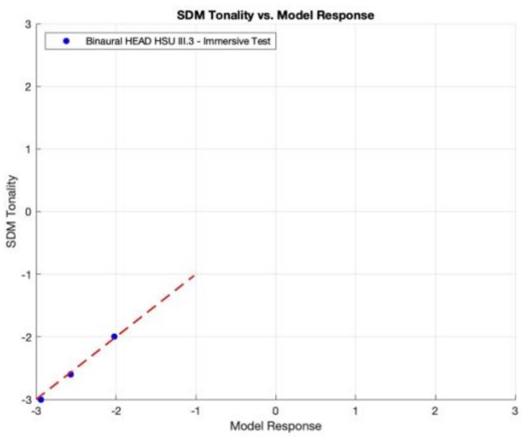


Fig 178-179 Comparison of subjective responses for the SDM fluctuation strength and the prediction model responses by Eq. (59) and Eq. (60). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Tonality**, relative to **Table 51**, revealed a hierarchy of correlation strength, with Loudness demonstrating the highest correlation at 0.99, followed by LZeq at 0.97. After some tests, the best prediction model for noise tonality relied on Loudness, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.07. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (61) has been determined.
The analysis of the correlation reported in **table 52**, involving **Tonality** and the NTi values, revealed a hierarchy of correlation strength, with LZeq demonstrating the highest correlation at 0.99, followed by Sharpness at 0.98. After some tests, the best prediction model for noise tonality relied on LZeq and Sharpness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.0009. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (62) has been determined. **Figure 180** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 181** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Tonality = \ -3.114\ +\ 0.038\ \times\ Loudness \tag{61}$$

$$SDM\ Tonality = \ -3.215 + 0.063\ \times\ Loudness +\ 2.049\ \times\ Tonality \tag{62}$$



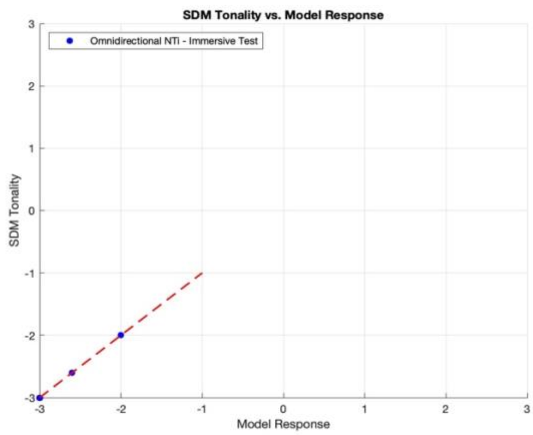


Fig 180-181 Comparison of subjective responses for the SDM Tonality and the prediction model responses by Eq. (61) and Eq. (62). The straight red line indicates the cases where model responses are equal to subjective responses.

3.3.6 Compared Tests – Immersive Listening Test with Headphones

Table 57 and **58** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	2	-2	-1.4	-1.6	-1.6	-2.4
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	2.8	-0.6	-0.6	-1.2	-2.4	-2.4
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	6.6	1.6	0.8	-0.2	-1.4	-0.6
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	3.2	-0.6	-0.6	-0.8	-2.4	-1.8

Table 57 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	2	-2	-1.4	-1.6	-1.6	-2.4
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	2.8	-0.6	-0.6	-1.2	-2.4	-2.4
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	6.6	1.6	0.8	-0.2	-1.4	-0.6
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	3.2	-0.6	-0.6	-0.8	-2.4	-1.8

Table 58 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.6.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 59**. In **Table 60** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.99307259	0.92730126	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.55091071	0.71864265	0.45897171	0.47157331	1								
Tonality	0.94444507	0.86246963	0.95647201	0.88631782	0.50729615	1							
Fluctuation Strength	0.9262928	0.96966181	0.88882791	0.95377929	0.61011667	0.75566028	1						
Annoyance Scale	0.39683417	0.57675042	0.31409297	0.48250205	0.50375089	0.08977216	0.71286008	1					
SDM Loudness	0.46155605	0.66144719	0.36607122	0.51123955	0.69230134	0.19293747	0.7458588	0.97177373	1				
SDM Roughness	0.44398462	0.64371338	0.3493051	0.49882624	0.66753488	0.16882139	0.73591291	0.97887258	0.99939981	1			
SDM Sharpness	0.67028694	0.81456148	0.59530534	0.72424355	0.65761614	0.41204802	0.89687952	0.94355833	0.96042628	0.9580119	1		
SDM Fluctuation Strength	-0.333625	-0.2734841	-0.3423578	-0.1758426	-0.4567422	-0.6002606	-0.0322217	0.50911405	0.30669692	0.33948467	0.24406017	1	
SDM Tonality	0.50157861	0.63449568	0.43850936	0.60247454	0.40008811	0.19158998	0.78689141	0.97582105	0.91890016	0.92766253	0.94720445	0.53773286	1

Table 59 Correlation coefficients for the Compare test – immersive listening test with headphones. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.96061386	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.95816669	0.84688846	0.83220306	1								
Tonality	0.98274366	0.9379216	0.9577757	0.91049069	0.87514514	1							
Fluctuation Strength	0.94064849	0.98089541	0.95354956	0.97454692	0.90837314	0.8749913	1						
Annoyance Scale	0.38740686	0.53948867	0.40817508	0.51637547	0.54074386	0.21682829	0.65445021	1					
SDM Loudness	0.49386724	0.63104039	0.45938927	0.545355	0.68251439	0.32470071	0.70560454	0.97177373	1				
SDM Roughness	0.47162858	0.61209263	0.44351334	0.53326976	0.65911715	0.30085676	0.69264242	0.97887258	0.99939981	1			
SDM Sharpness	0.67082027	0.78783927	0.674447	0.7513442	0.77446699	0.52706426	0.86345823	0.94355833	0.96042628	0.9580119	1		
SDM Fluctuation Strength	-0.4514196	-0.3166657	-0.2877394	-0.1558032	-0.4222842	-0.5501993	-0.1281056	0.50911405	0.30669692	0.33948467	0.24406017	1	
SDM Tonality	0.45548086	0.59761842	0.52206546	0.63111144	0.53716843	0.30378559	0.72691973	0.97582105	0.91890016	0.92766253	0.94720445	0.53773286	1

Table 60 Correlation coefficients for the Compare test - immersive listening test with headphones. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 59**, the analysis revealed a hierarchy of correlation strength too low to be considered suitable for prediction.

The analysis of the data reported in **table 60**, revealed a hierarchy of correlation strength too low to be suitable for the prediction model.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 61 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 62** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	2
V4	33.5	32.5	35	32.5	2.8
V8	50	50	52	47.5	6.6
V12	59	60	62	57	3.2

Table 61 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99977304	1			
PNC	0.99815903	0.99920904	1		
RNC	0.99660463	0.99498405	0.99037924	1	
Annoyance Scale	0.56620186	0.56337375	0.56520895	0.53188844	1

Table 62 Correlation coefficients for the Compare test - immersive listening test with headphones. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the subjective and objective parameters is too low to be suitable for the linear regression.

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 63** shows the data used for the prediction and **table 64** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	2
V4	25	24	23	25	2.8
V8	42	41	42	40	6.6
V12	50	52	51	48	3.2

Table 63 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.99836754	1		
RNC	0.99975084	0.99938348	0.99975183	1	
Annoyance Scale	0.5873036	0.54405217	0.58669327	0.57131267	1

Table 64 Correlation coefficients for the Compare test - immersive listening test with headphones. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

The correlation between the subjective and objective parameters is too low to be suitable for the linear regression.

3.3.7.2 Semantic Differential Method

As shown in **table 59** and **table 60**, the correlation coefficients calculated between the subjective and objective parameters is not suitable for linear regression.

3.3.7 Compared Tests - Binaural Test

Table 65 and **66** show the mean values of subjective assessment for each type of assessment method, along with the corresponding psychoacoustic parameters extracted from recordings captured by both HEAD HSU III.3 and omnidirectional microphone NTi. The psychoacoustic parameters are reported as the average between the left and right channels for each recording.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	63.85	27.84	1.17	0.623	0.065	0.0861	0.002925	1	-3	-3	-3	-3	-3
V4	67.39	43.61	4.52	0.734	0.1855	0.1685	0.004675	1.4	-2.6	-1.4	-2.4	-1.4	-2
V8	73.38	59.58	14.2	1.45	0.172	0.179	0.00989	3.8	-0.6	-0.4	-1.8	-0.6	-2.2
V12	80.49	67.86	28.4	1.895	0.1635	0.385	0.0101	5	0.6	0.6	-1.8	-1.2	-1.4

Table 65 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Background	56.53	23.73	0.864	0.524	0.0658	0.078	0.00241	1	-3	-3	-3	-3	-3
V4	65.8	41.8	4.62	0.663	0.131	0.157	0.00466	1.4	-2.6	-1.4	-2.4	-1.4	-2
V8	70.7	57.5	14.6	1.44	0.147	0.181	0.00926	3.8	-0.6	-0.4	-1.8	-0.6	-2.2
V12	80.2	68.6	24.7	1.85	0.163	0.315	0.0102	5	0.6	0.6	-1.8	-1.2	-1.4

Table 66 Psychoacoustic parameters: LZeq, LAeq, Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance), Semantic Differential Method (SDM): Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength.

3.3.4.1 Rating Scale Method

Based on the results of the subjective evaluation using the 1-10 Annoyance Rating Scale, the independent variables of the model were selected by examining correlation coefficients between the psychoacoustic parameters recorded with HEAD HSU III.3 and the rating scale, as presented in **Table 67**. In **Table 68** are displayed the correlation coefficients calculated with the recordings from the omnidirectional microphone NTi.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.96475961	1											
Loudness	0.99307259	0.927301255	1										
Sharpness	0.9862058	0.95164994	0.98115858	1									
Roughness	0.55091071	0.718642653	0.45897171	0.471573312	1								
Tonality	0.94444507	0.862469629	0.95647201	0.886317821	0.507296149	1							
Fluctuation Strength	0.9262928	0.969661811	0.88882791	0.953779286	0.610116675	0.75566028	1						
Annoyance Scale	0.9811735	0.957925152	0.97147141	0.99883405	0.488952665	0.86755923	0.96684007	1					
SDM Loudness	0.98721357	0.958725502	0.97935688	0.999699065	0.492086977	0.88513089	0.95949498	0.99932881	1				
SDM Roughness	0.96496673	0.994132475	0.92900612	0.934316998	0.748574302	0.89568246	0.93758465	0.93720823	0.94169711	1			
SDM Sharpness	0.8947137	0.979790467	0.83731734	0.891932478	0.783745381	0.74786241	0.97026636	0.90744251	0.90267093	0.9638576	1		
SDM Fluctuation Strength	0.68727097	0.853946477	0.59724567	0.672254438	0.910644257	0.52482371	0.8372913	0.69864636	0.69006556	0.84125086	0.93435318	1	
SDM Tonality	0.86695046	0.876432119	0.83778088	0.776548688	0.79305243	0.91954768	0.73217427	0.76782574	0.78471401	0.92333854	0.81663822	0.72364641	1

Table 67 Correlation coefficients for the Compared tests - binaural test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

	Lzeq	Laeq	Loudness	Sharpness	Roughness	Tonality	Fluctuation Strength	Annoyance Scale	SDM Loudness	SDM Roughness	SDM Sharpness	SDM Fluctuation Strength	SDM Tonality
Lzeq	1												
Laeq	0.9852396	1											
Loudness	0.96821564	0.960613856	1										
Sharpness	0.94089113	0.95585422	0.99015344	1									
Roughness	0.9398703	0.958166689	0.84688846	0.832203062	1								
Tonality	0.98274366	0.937921602	0.9577757	0.910490685	0.875145144	1							
Fluctuation Strength	0.94064849	0.980895409	0.95354956	0.974546922	0.908373138	0.874991297	1						
Annoyance Scale	0.93790881	0.954577559	0.9885797	0.999941287	0.830145532	0.906042645	0.975277253	1					
SDM Loudness	0.94647028	0.957091483	0.99341246	0.999663627	0.83412162	0.920127487	0.971342527	0.999328814	1				
SDM Roughness	0.99061596	0.997824093	0.95196751	0.939110649	0.969036639	0.948311847	0.966089804	0.937208229	0.94169711	1			
SDM Sharpness	0.9194853	0.971826906	0.88572703	0.906746832	0.959160663	0.831193926	0.976908211	0.907442511	0.90267093	0.9638576	1		
SDM Fluctuation Strength	0.76267138	0.841906172	0.6670237	0.697199094	0.922432354	0.635993698	0.838651107	0.698646363	0.69006556	0.84125086	0.93435318	1	
SDM Tonality	0.94230345	0.896462758	0.8363921	0.77349816	0.929859284	0.944239334	0.79313633	0.76782574	0.78471401	0.92333854	0.81663822	0.72364641	1

Table 68 Correlation coefficients for the Compared tests - binaural test. Both objective (omnidirectional microphone NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale, while SDM: Loudness, Sharpness, Roughness, Tonality, Fluctuation Strength correspond to the Semantic Differential Method (SDM) scales.

Starting with the data reported in **table 67**, the analysis revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.99, followed by LZeq at 0.98. After some tests, the best prediction model for noise annoyance relied on LAeq and Roughness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.03. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (63) has been determined. **Figure 182** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -1.846 + 0.136 \times LAeq - 14.365 \times Roughness \qquad (63)$$

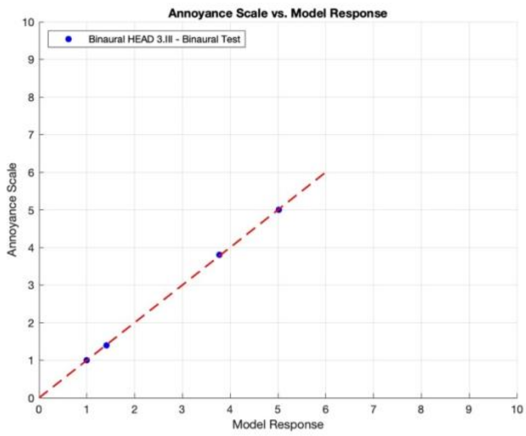


Fig 182 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (63). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis of the data reported in **table 68**, revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.99, followed by LZeq at 0.99. After some tests, the best prediction model for noise annoyance relied on Sharpness, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.24. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (64) has been determined. **Figure 183** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$Annoyance = -0.86 + 3.876 \times Sharpness \quad (64)$$

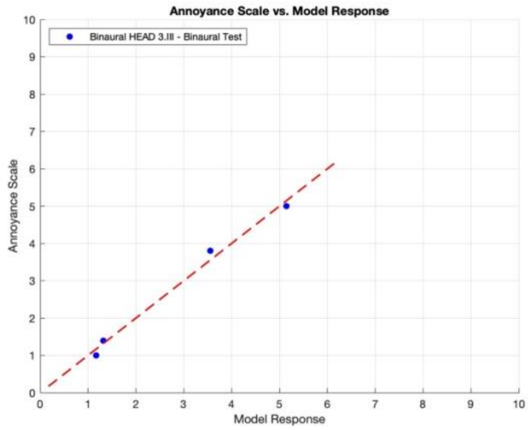


Fig 183 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (64). The straight red line indicates the cases where model responses are equal to subjective responses.

Following the article written by Antti Kuusinen, the linear regression of the Annoyance Rating Scale is calculated with parameters based on the spectrum form, analyzed in the previous chapters, such as Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion, calculated with both artificial head and omnidirectional microphone.
Table 69 shows the data used for the prediction, using objective recordings from the artificial head HSU III.3, and **table 70** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	19	15.5	15.5	22.5	1
V4	33.5	32.5	35	32.5	1.4
V8	50	50	52	47.5	3.8
V12	59	60	62	57	5

Table 69 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with HEAD HSU III.3; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99977304	1			
PNC	0.99815903	0.999209045	1		
RNC	0.99660463	0.994984048	0.99037924	1	
Annoyance Scale	0.96698228	0.961524409	0.94983992	0.983137988	1

Table 70 Correlation coefficients for the Compared tests - binaural test. Both objective (HEAD HSU III.3) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

After some tests, the best prediction model for noise annoyance relied on Noise Criterion and Preferred Noise Criterion, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.007. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (65) has been determined. **Figure 184** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response. The values of RMSE and the adjuster R-Square are better than the ones reported above, showing that the parameters based on spectral shape are a better indicator of the annoyance prediction model.

$$Annoyance = -0.845 + 0.767 \times NC - 0.65 \times PNC \quad (65)$$

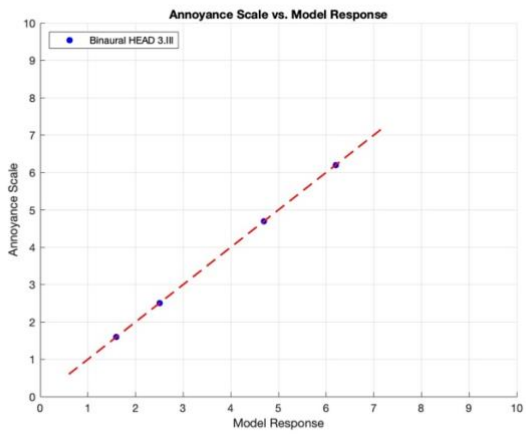


Fig 184 Comparison of subjective responses for the 1-10 Annoyance Rating Scale and the prediction model responses by Eq. (65). The straight red line indicates the cases where model responses are equal to subjective responses

The same methodology is used for the data recorded with the omnidirectional microphone NTi. **Table 71** shows the data used for the prediction and **table 72** reports the correlation coefficients.

	NR	NC	PNC	RNC	Annoyance Scale
Background	9	8	5	11	1
V4	25	24	23	25	1.4
V8	42	41	42	40	3.8
V12	50	52	51	48	5

Table 71 Psychoacoustic parameters: Room Noise Criterion, Noise Criterion, Noise Rating and Preferred Noise Criterion recorded with NTi; Subjective ratings: Annoyance Scale (1-10 Rating Scale Annoyance).

	NR	NC	PNC	RNC	Annoyance Scale
NR	1				
NC	0.99835716	1			
PNC	0.99999901	0.998367538	1		
RNC	0.99975084	0.999383483	0.99975183	1	
Annoyance Scale	0.79367219	0.805945404	0.79449976	0.797771164	1

Table 72 Correlation coefficients for the Compared tests - binaural test. Both objective (NTi) and subjective data are reported. The 1-10 Rating Scale corresponds to the Annoyance Rating Scale.

As shown in **Table 72**, the correlation between the subjective and objective parameters is not high enough to allow a reliable prediction model.

3.3.4.2 Semantic Differential Method

Similarly to the analysis made in the previous paragraph, linear regressions were calculated as well for the results collected with the semantic differential method scale. The results are correlated with both the parameters extracted from the recordings made with the artificial head HEAD HSU III.3 and the omnidirectional microphone NTi, in order to see which one gives the most reliable results.
In **Table 67** are displayed the correlation coefficients calculated with the recordings from the artificial head HEAD HSU III.3 and in **table 68** the ones from the omnidirectional microphone NTi.
Starting with the data reported in **table 67**, the analysis for the subjective parameter **Loudness** revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.99, followed by LZeq at 0.987. After some tests, the best prediction model for noise loudness relied on Sharpness, resulting in

an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.05. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (66) has been determined.
The analysis of the correlation reported in **table 68**, involving **Loudness** and the NTi values, revealed a hierarchy of correlation strength, with Sharpness demonstrating the highest correlation at 0.996, followed by LZeq at 0.994. After some tests, the best prediction model for noise loudness relied on Sharpness, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.17. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (67) has been determined. **Figure 185** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 186** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Loudness = -4.704 + 2.81 \times Sharpness \quad (66)$$

$$SDM\ Loudness = -4.643 + 3.435 \times Sharpness \quad (67)$$

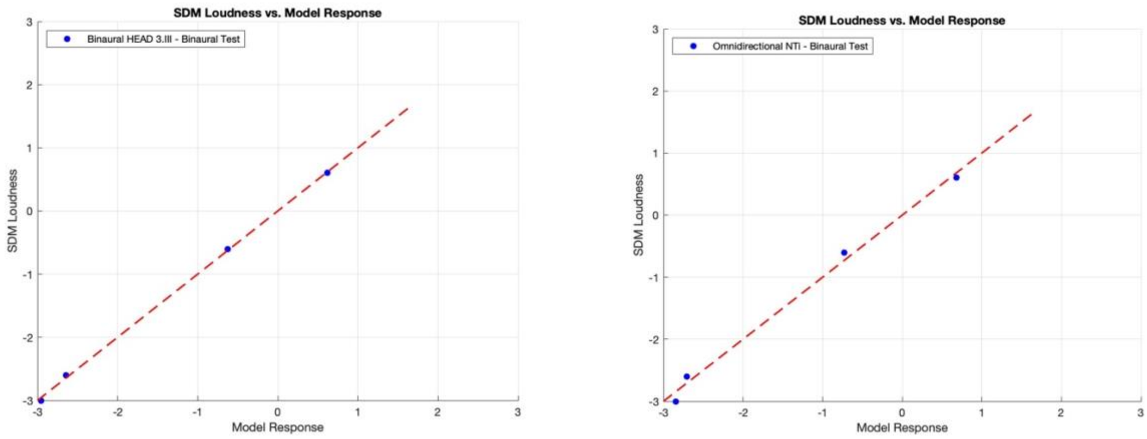


Fig 185-186 Comparison of subjective responses for the SDM Loudness and the prediction model responses by Eq. (66) Eq. (67). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Roughness**, relative to **Table 67**, revealed a hierarchy of correlation strength, with LAeq demonstrating the highest correlation at 0.99, followed by LZeq at 0.96. After some tests, the best prediction model for noise roughness relied on LAeq, resulting in an **adjusted R-square** value of 0.98 and a Root Mean Square Error (**RMSE**) of 0.20. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (68) has been determined.
The analysis of the correlation reported in **table 68**, involving **Roughness** and the NTi values, revealed a hierarchy of correlation strength, with Fluctuation strength demonstrating the highest correlation at 0.99, followed by LAeq at 0.98. After some tests, the best prediction model for noise roughness relied on Roughness and Fluctuation strength, resulting in an **adjusted R-square** value of 0.99 and a Root Mean Square Error (**RMSE**) of 0.015. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (69) has been determined. **Figure 187** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 188** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ Roughness = -5.332 + 0.086 \times LAeq \quad (68)$$

$$SDM\ Roughness = -5.303 + 19.8 \times Roughness + 414.274 \times Fluctuation\ Strength \quad (69)$$

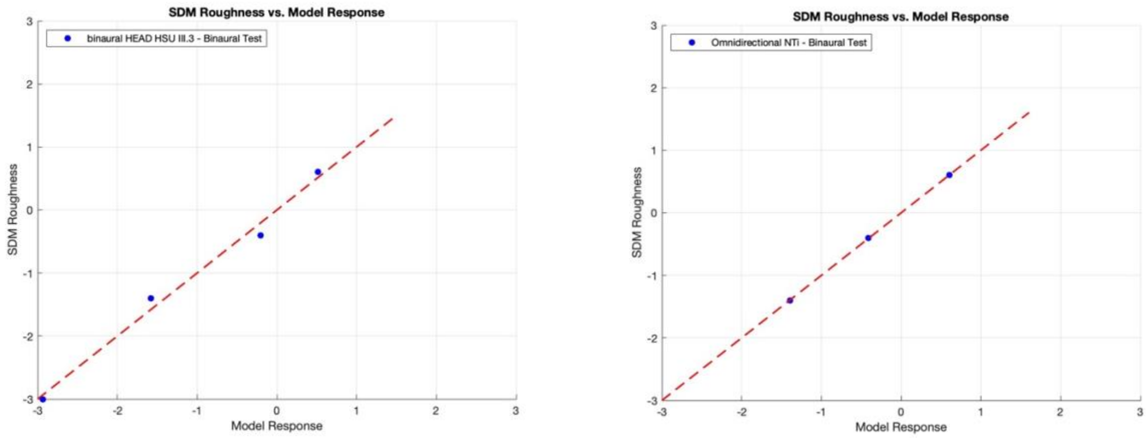


Fig 191-192 Comparison of subjective responses for the SDM Roughness and the prediction model responses by Eq. (68) and Eq. (69). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Sharpness**, relative to **Table 67**, revealed a hierarchy of correlation strength with LAeq demonstrating the highest correlation at 0.98, followed by Fluctuation Strength at 0.97. After some tests, the best prediction model for sharpness relied on LAeq, resulting in an **adjusted R-square** value of 0.93 and a Root Mean Square Error (**RMSE**) of 0.14. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (70) has been determined.

The analysis for the subjective parameter **Sharpness**, relative to **Table 68**, revealed a hierarchy of correlation strength with Roughness demonstrating the highest correlation at 0.96, followed by LAeq at 0.96. After some tests, the best prediction model for sharpness relied on LAeq and Loudness, resulting in an **adjusted R-square** value of 0.999 and a Root Mean Square Error (**RMSE**) of 0.0009. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (71) has been determined. **Figure 189** illustrates the subjective response of the test versus the predicted values of the equation. **Figure 190** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Sharpness = \ -3.829\ +\ 0.032\ \times LAeq \qquad (70)$$

$$SDM\ \ Sharpness = \ -4.657 + 0.074\ \times\ LAeq\ - 0.112\ \times Loudness \qquad (71)$$

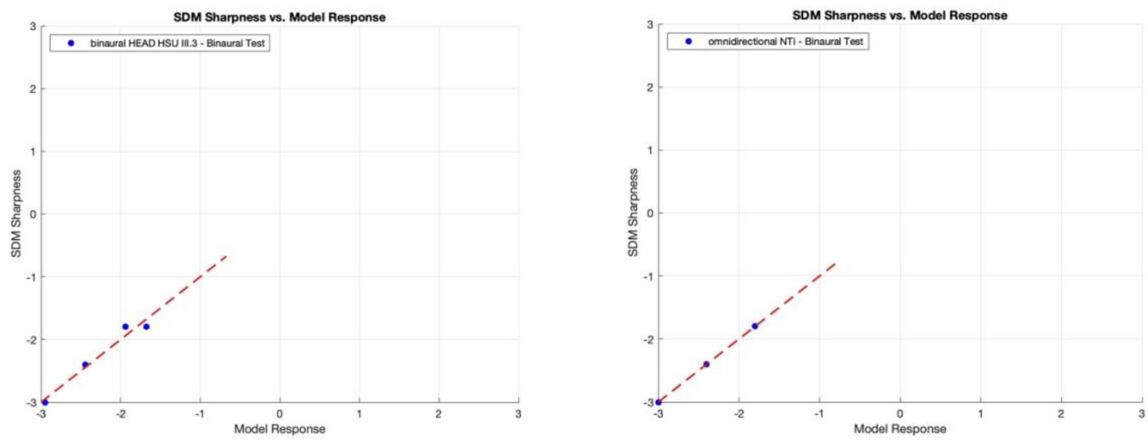


Fig 189-190 Comparison of subjective responses for the SDM Sharpness and the prediction model responses by Eq. (70) Eq. (71). The straight red line indicates the cases where model responses are equal to subjective responses.

The analysis for the subjective parameter **Fluctuation Strength**, relative to **Table 67**, revealed a hierarchy of correlation strength too low to be used for prediction models.
The analysis of the correlation reported in **table 68**, involving **Fluctuation Strength** and the NTi values, revealed a hierarchy of correlation strength, with Roughness demonstrating the highest correlation at 0.96. After some tests, the best prediction model for fluctuation strength relied on Roughness, resulting in an **adjusted R-square** value of 0.88 and a Root Mean Square Error (**RMSE**) of 0.36. Importantly, all p-values were below 0.05, indicating the statistical significance of the findings. The multiple linear regression equation (72) has been determined. **Figure 191** illustrates the subjective response of the test versus the predicted values of the equation. The results closely align with the red line, indicating that the subjective response is in line with the model response.

$$SDM\ \ Fluctuation\,Strength = \ -484 + 28.314\ \times\ Roughness \tag{72}$$

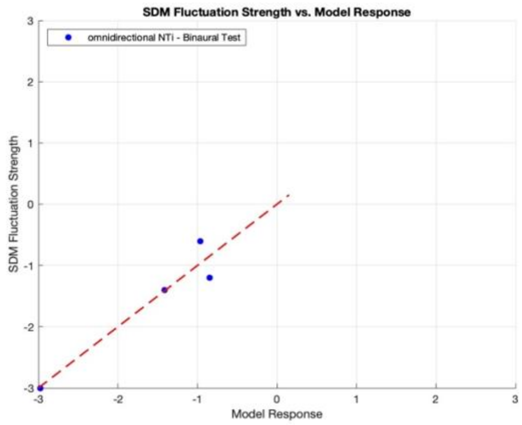


Fig 191 Comparison of subjective responses for the SDM fluctuation strength and the prediction model responses by Eq. (72). The straight red line indicates the cases where model responses are equal to subjective responses.

4. Discussion

This section analyses the ability of different tests to predict sound quality across various metrics.
All R-squared values of multiple linear regressions were higher than 0.88, with a maximum of 0.999, indicating strong predictive capability across the evaluated perceptual dimensions.
Below is the analysis and comparison of the results obtained in the laboratory and those calculated using data from the artificial head HSU III.3 recordings with the real-case scenario. **Table 73-79** show the comparisons.
Regarding the parameter of Annoyance, in comparison to the results predicted using the data from the tractor test, we observe similarities between the immersive test and the binaural test. Both use the LAeq/LZeq parameter for prediction but are not identical, as they employ additional parameters.
For the regression calculated for Annoyance using spectrum-based parameters, the equations calculated with the immersive test data use the same parameter, and while the coefficients differ, the deviation is negligible.
The equation related to Loudness does not show similarity to the real-case scenario, but in the laboratory tests, Tonality and Fluctuation Strength are recurring parameters.
The predictive equation for Roughness shows a strong similarity with the one calculated in the immersive test, but no other similarities are observed with the other laboratory tests.

The predictive equation for Fluctuation Strength shows an affinity with the equation obtained from the immersive test. The equations calculated with other subjective tests share only one of the two parameters, and the coefficients are significantly different.
For Sharpness, there is no predictive equation derived from the field test; thus, comparison with the real case is not possible, and the laboratory equations are significantly different.
The equation related to Tonality cannot be compared with the real-case scenario, and the laboratory equations are significantly different.

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Annoyance = -0.633 + 0.098 × Laeq	0.92	0.49
Immersive Test	Annoyance − 0.64 + 0.075×$LAeq$ − 152.43×$Fluctuation\ Strength$	0.99	0.01
Binaural Test	Annoyance = −22.121 + 0.379 × $LZeq$ − 5.769 × $Tonality$	0.99	0.02
Immersive headphone Test	Annoyance = 2.97+722.24×$Fluctuation\ Strength$-3.41×$Sharpness$	0.99	0.07
Favaretto	Annoyance = −3.9563 + 0.0956 ×$LZeq$ + 0.1660 ×$Loudness$	0.97	0.49

Table 73 Annoyance Linear regression using HSU III.3 values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Annoyance = 0.710 + 0.086 × PNC	0.94	0.43
Immersive Test	Annoyance = 0.434 + 0.039 × PNC	0.97	0.14
Binaural Test	Annoyance = −2.826 −0.105 × NC + 0.269 × RNC	0.99	0.003
Immersive headphone Test	/	/	/

Table 74 Annoyance Linear regression using HSU III.3 values and spectrum-based parameters

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Loudness = -12.953 + 0.181 × $LZeq$	0.97	0.22
Immersive Test	Loudness = −3.694 + 213.035. × $Fluctuation\ Strength$	0.94	0.19
Binaural Test	Loudness = −2.833 −3.144 × $Tonality$ + 2.553 × $Sharpness$	0.99	0.02
Immersive headphone Test	Loudness =-2.16 -4.20× $Tonality$ + 314.84 × $Fluctuation\ Strength$	0.99	0.003
Favaretto	Loud = −6.1938 + 0.0613 ×$LZeq$ + 0.1084 ×$Loudness$	0.97	0.33

Table 75 Loudness Linear regression using HSU III.3 values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Roughness = -2.129 + 0.033 × $LAeq$	0.94	0.15
Immersive Test	Roughness = −5.379 + 0.084 × $LAeq$	0.98	0.21
Binaural Test	Roughness = −1.856 + 0.081 × $Loudness$	0.99	0.09
Immersive headphone Test	Roughness = 0.309-3.040 x $Sharpness$ + 618.22 $Fluctuation\ Strength$	0.99	0.003
Favaretto	Rough = −9.1139 + 0.1548 ×$LZeq$ − 12.3291 × $Roughness$	0.88	0.38

Table 76 Roughness Linear regression using HSU III.3 values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	/	/	/
Immersive Test	$Sharpness = -3.694 + 213.035 \times Fluctuation\ Strength$	0.94	0.19
Binaural Test	/	/	/
Immersive headphone Test	$Sharpness = 0.390 + 3.502 \times Tonality - 10.629 \times Roughness$	0.99	0.01

Table 77 Sharpness Linear regression using HSU III.3 values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Fluctuation S. = -6.594 + -1.991 × *Roughness* + 0.096 × *LZeq*	0.99	0.001
Immersive Test	Fluctuation S. = −4.465 + 12.184 × *Roughness* + 0.028 × *LZeq*	0.99	0.01
Binaural Test	Fluctuation S. = −2.516 + 177.226×*FluctuationS.* −0.003×*Loudness*	0.99	0.001
Immersive headphone Test	Fluctuation S. = 0.390 + 3.502 × *Tonality* -10.629 × *Roughness*	0.99	0.01

Table 78 Fluctuation Strength Linear regression using HSU III.3 values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	/	/	/
Immersive Test	$Tonality = -3.498 + 0.743 \times Sharpness$	0.99	0.04
Binaural Test	/	/	/
Immersive headphone Test	$Tonality = 89.602 + 0.793 \times Loudness - 1.403 \times Lzeq$	0.99	0.09

Table 79 Tonality Linear regression using HSU III.3 values

Below is the analysis and comparison of the results obtained in the laboratory and those calculated using data from the NTi omnidirectional microphone recordings with the real-case scenario. **Table 80-86** show the comparisons.

Regarding the parameter of Annoyance, in comparison to the results predicted using data from the tractor test, the only similarity is that every test uses LZeq for prediction. However, the immersive test and the immersive test with headphones test yielded very similar results, using the same parameter but with different coefficients. For the regression calculated for Annoyance using spectrum-based parameters, no similarities are observed with the real-case scenario. However, the immersive test and the binaural test yielded very similar results, again using the same parameter but with different coefficients.

The equation related to Loudness shows a similarity between the immersive test and the tractor test, as both are based on the LAeq parameter but with different coefficients.

The predictive equation for Roughness shows a partial similarity with the immersive headphone listening test, as they share a common parameter.

The predictive equation for Fluctuation Strength shows no affinity with the real-case scenario or with the laboratory tests conducted, except for the binaural test that shares a common parameter.

For Sharpness, comparison with the real-case scenario is not possible.

The equation related to Tonality cannot be compared with the real-case scenario.

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Annoyance = -3.722+26.921×Roughness+0.66×LZeq	0.99	0.04
Immersive Test	Annoyance = 20.55 + 0.48 ×*LZeq* -44.12 ×*Tonality*	0.99	0.01
Binaural Test	Annoyance = -0.39 + 0.03 ×*Laeq* + 2.37×*Sharpness*	0.99	0.13
Immersive headphone Test	Annoyance = 20.55+0.48 ×LZeq -44.12 ×Tonality	0.99	0.1
Favaretto	Annoyance = −4.4705 + 0.1043 ×*LZeq* + 0.0844 ×*Loudness*	0.97	0.35

Table 80 Annoyance Linear regression using NTi values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Annoyance = 75.42 -16.3 × NR + 14.62 × NC	0.99	0.06
Immersive Test	Annoyance = 0.752 + 0.041 × *PNC*	0.96	0.16
Binaural Test	Annoyance = *0.422* + *0.106* × *PNC*	0.95	0.16
Immersive headphone Test	/	/	/

Table 81 Annoyance Linear regression using NTi values and spectrum-based parameters

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Loudness = -2.41+0.06×Loudness-0.035×Laeq	0.99	0.1
Immersive Test	Loudness = 0.08 + 0.04×*Laeq*	0.97	0.12
Binaural Test	Loudness = *-2.742* + *2.400 x Sharpness*	0.95	0.26
Immersive headphone Test	/	/	/
Favaretto	Loudness = − 5.9836 + 0.0626×*LZeq* + 0.0776×*Loudness*	0.97	0.25

Table 82 Loudness Linear regression using NTi values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Roughness = -2.17+11.73×Roughness+1.14×Tonality	0.99	0.06
Immersive Test	Roughness = -5.186 + 0.098 *x Laeq*	0.97	0.26
Binaural Test	Roughness = *-1.9* + *0.09*× *Loudness*	0.95	0.21
Immersive headphone Test	Roughness = -19.3 + 0.396 × Lzeq -37.12 ×Tonality	0.99	0.04
Favaretto	Rough = − 4.9950 + 3.5662 ×*Roughness* + 0.0529 ×*LZeq*	0.88	0.27

Table 83 Roughness Linear regression using NTi values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	/	/	/
Immersive Test	/	/	/
Binaural Test	/	/	/
Immersive headphone Test	Sharpness = *−6.02* + *−36.06* ×*Roughness* + *0.14* × *Lzeq*	0.99	0.04

Table 84 Sharpness Linear regression using NTi values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	Fluctuation S. = -0.388+0.081×Loudness-0.012×Laeq	0.99	0.01
Immersive Test	Fluctuation S. = -4.902 + 30.995 *x Roughness*	0.99	0.12
Binaural Test	Fluctuation S.-2.53 -0.016×Loudness+209.5×Fluctuation Strength	0.98	0.08
Immersive headphone Test	/	/	/

Table 85 Fluctuation Strength Linear regression using NTi values

Rating Scale	Regression Equation	R-squared	RMSE
On tractor Test	/	/	/
Immersive Test	Tonality = 3.42 + 0.71 × *Sharpness*	0.99	0.05
Binaural Test	/	/	/
Immersive headphone Test	Tonality = -6.26 + -70.53 × *Roughness* + 0.21 ×*LZeq*	0.92	0.36

Table 86 Tonality Linear regression using NTi values

Thus, the immersive test has proven to be the most accurate in terms of prediction compared to the tractor tests taken as a reference. This suggests that the immersive test may provide a more reliable estimation of sound perception, particularly in environments where Roughness plays a significant role. The immersive test has a notable advantage in accurately reproducing low frequencies, making it particularly relevant in scenarios involving low-frequency noise sources such as tractor engines.

Moreover, we can conclude that the artificial head produced by Head Acoustics is indeed a tool that offers greater reliability in terms of prediction quality.

As for the results obtained between the compared tests and the actual laboratory tests, it can be said that the immersive headphone listening test did not yield results. The immersive test was quite accurate in parameter comparison, while the binaural test showed similarities but to a lesser extent than the immersive test. Regressions performed with NTi values produced significantly different results.

The prediction results obtained using parameters related to the spectral shape of the sound proved to be very accurate and provided better outcomes in most of the calculations performed.

There are similarities between the prediction model calculated with the artificial head and the omnidirectional microphone.

The introduction of Fluctuation Strength and Tonality parameters into the methodology brought improvements and was very useful in predicting parameters such as Fluctuation Strength, Sharpness, and Annoyance.

In comparison to Favaretto's research [39], these models are more accurate in terms of accuracy, the value of R-squared and RMSE are improved, thanks to the use of parameters like fluctuation strength and tonality and the microphones used, especially the artificial head HSU III.3 that has better qualities in terms of impulse response and dynamic range.

Our prediction models outperformed two out of three previous studies that used similar methodologies, particularly in predicting noise Annoyance and Roughness. This suggests that our approach with spatialized audio offers advancements in sound perception assessment. However, to further increase prediction reliability, future research should consider testing a higher number of HVAC operating modes and conducting in-field tests with tractors moving under different scenarios.

5. Conclusions

An extensive review of the current literature regarding HVAC noise within tractor cabins was undertaken. This will serve as a baseline in order to identify the most significant attributes that characterize HVAC noise perception and the prevailing techniques in sound quality engineering used to evaluate this type of noise for tractors. The primary psychoacoustic parameters identified as significant contributors to HVAC noise annoyance and sound quality include Linear Sound Pressure Level (SPL), A-weighted Sound Pressure Level (A-SPL), Loudness, Roughness, Fluctuation Strength, Tonality and Sharpness. The thesis explores two subjective evaluation methods, including the 1-10 Rating Scale Method and the Semantic Differential Method. These were tested through binaural listening, immersive and spatialized audio in Audio Space Lab, as Marco Favaretto did in his study [39] and a novel playback method: the immersive listening test with spatialized audio of 3rd order ambisonics played in headphones. This new method allowed us to compute linear regressions with impressive RMSE and R-squared results, but the comparison with the test done on the tractor are not brilliant.

Through objective and subjective data, a predictive model was developed for Annoyance, Loudness, Roughness, Sharpness, Tonality and Fluctuation Strength using multiple linear regression. The study aimed to compare the effectiveness of the Immersive Test against the Binaural Test and the immersive listening test with headphones in forecasting sound perception across various metrics. Our findings suggest that the Immersive Test exhibits slightly superior predictive performance compared to the other tests. Furthermore, the study aimed to find the best reproductive audio system in order to have the most similar results as the tests conducted on the tractor. In this case, the immersive test is the one that gave the best results. In addition, the most accurate microphone in terms of accuracy of prediction results, the artificial head made by Head Acoustics is more reliable.

In conclusion, according to Favaretto's study [39], the most reliable playback system is the immersive test.

6. Acknowledgements

I would like to start by thanking Professor Arianna Astolfi for teaching me so much, guiding me in this research, and pushing me to always do better. I would also like to thank Professor Louena Shtrepi for her advice, knowledge, and words of encouragement; her help has been absolutely invaluable. I would like to thank Fabrizio Ferrian from HeadAcoustics as well, for his availability, professionalism, and immense help with the measurements and subsequent analysis. A special thanks also goes to Angela Guastamacchia for her support, her boundless willingness to help, and her encouragement. Thanks also to Marco Favaretto for all the times I reached out to him for advice and for his precious work.

Finally, I would like to thank my parents, my friends, and Samuele, who supported, encouraged, and put up with me during this great adventure.

7. References

1. Ricardo Penna Leite; Stephan Paul; Samir N.Y. Gerges "A sound quality-based investigation of the HVAC system noise of an automobile model"- Applied Acoustics (2009)
2. Stephen Grigg et all "Acoustic Sensing and Noise Identification of a Heating, Ventilation and Air Conditioning Unit: Industrial Case Study" - Applied science (2021)
3. Nabil Hammad "Influence of Air Conditioning Characteristics on Vehicle Interior Noise" - International Journal of Vehicle Structures & Systems (2010)
4. Ji-Hyun Yoon et all "Reliability improvement of a sound quality index for a vehicle HVAC system using a regression and neural network model" - Applied Acoustics (2012)

5. Jiseon Back et all "Design and implementation of comfort-quality HVAC sound inside a vehicle cabin" - Applied Acoustics (2021)
6. Sneha Singh; A.R. Mohanty "HVAC Noise control using natural materials to improve vehicle interior sound quality" - Applied Acoustic (2018)
7. M. Satya Prasad et all "Rating criteria for vehicle in-cabin noise with emphasis on HVAC system operation" - International Journal of Vehicle Noise and Vibration (2009)
8. Antti Kuusinen; Valtteri Hongisto "Best descriptors for predicting noise annoyance and spectrum-based penalty of steady-state wideband noise" – Internoise (2024)
9. Manuela Gragnaniello; Claudia Celletti; Alessandra Resca; Giovanni Galeoto; Filippo Camerota "Italian Version of the YQOL-DHH Questionnaire: Translation and cross-cultural adaptation" – American Academy of Otolaryngology (2021)
10. Ma C.; Chen C.; Liu Q.; Gao H.; Li Q.; Gao H.; Shen Y. "Sound Quality Evaluation of the Interior Noise of Pure Electric Vehicle Based on Neural Network Model" - IEEE Transactions on Industrial Electronics (2017)
11. Zhipeng Wang; Yanyan Zuo; Liming Sun "The Impact of Sound Pressure Level, Loudness, Roughness, Sharpness, Articulation Index, Hand Vibration, and Seat Vibration on Subjective Comfort Perception of Tractor Drivers" - Symmetry,(2023)
12. Chen P.; Xu L.; Tang Q.; Shang L.; Liu W. "Research on prediction model of tractor sound quality based on genetic algorithm" - Applied Acoustics (2022)
13. Gao Y.-H.; Qian K.; Liang J.; LiuQ.; Zhao J. 1915. "Interior sound quality evaluation model of heavy commercial vehicles" - Journal of Vibroengineering (2016)
14. Yoon J.-H.; Yang I.-H.; Jeong J.-E.; Park S.-G.; Oh J.-E. "Reliability improvement of a sound quality index for a vehicle HVAC system using a regression and neural network model" - Applied Acoustics (2012)
15. Back J.; Lee S.-K.; Min Lee S.; An K.; Kwon D.-H.; Park D.-C. "Design and implementation of comfort-quality HVAC sound inside a vehicle cabin" - Applied Acoustics (2021)
16. PDISO/TS15666:2021
17. J1441_201609 "Subjective Rating Scale for Vehicle Ride and Handling" – SAE (2016)
18. Norm O.; Scott A; Chris E.; Scott L. "Guidelines for Jury Evaluation of Automotive Sound" – SAE (1999)
19. Yu. H.; Dou L. "An empirical category-ratio scale for evaluating the subjective intensity of noise based on the comparison of estimated magnitudes and categories" - Applied Acoustics (2020)
20. Jun Z.; Kean C.; Hao Li; Xingshu C.; Ningjuan D. "The effects of rating scales and individual characteristics on perceived annoyance in laboratory listening tests" - Applied Acoustics (2023)
21. Norm O.; Scott A.; Chris E.; Scott L. "Guidelines for Jury Evaluations of Automotive Sounds" – SAE Technical Papers (1999)
22. Liqiang Yanga; Pan Wanga; Jie Wang "Research on evaluation model for vehicle interior sound quality based on an optimized BiLSTM using genetic algorithm" - Mechanical Systems and Signal Processing (2023).
23. Wang Y.; Zhang S.; Meng D.; Zhang L. "Nonlinear overall annoyance level modeling and interior sound quality prediction for pure electric vehicle with extreme gradient boosting algorithm" - Applied Acoustics (2022)
24. Huang H.; Wu J.; Lim T.C.; Yang M.; Ding W. "Pure electric vehicle nonstationary interior sound quality prediction based on deep CNNs with an adaptable learning rate tree" - Mechanical Systems and Signal Processing (2021)
25. Huang X.; Huang H.; Wu J.; Yang M.; Ding W. "Sound quality prediction and improving of vehicle interior noise based on deep convolutional neural networks" - Expert Systems with Applications (2020)
26. Qian K.; Hou Z. "Intelligent evaluation of the interior sound quality of electric vehicles" - Applied Acoustics (2021)

27. Huang H.B.; Wu J.H.; Huang X.R.; Yang M.L.; Ding W.P. "The development of a deep neural network and its application to evaluating the interior sound quality of pure electric vehicles" - Mechanical Systems and Signal Processing (2019)
28. Li D.; Huang Y. "The discomfort model of the micro commercial vehicles interior noise based on the sound quality analyses" - Applied Acoustics (2018)
29. Wang Y. S.; Shen G. Q.; Xing Y. F. "A sound quality model for objective synthesis evaluation of vehicle interior noise based on artificial neural network" - Mechanical Systems and Signal Processing (2014)
30. Zhang E.; Hou L.; Shen C.; Shi Y.; Zhang Y. "Sound quality prediction of vehicle interior noise and mathematical modeling using a back propagation neural network (BPNN) based on particle swarm optimization (PSO)" - Measurement Science and Technology (2015)
31. Huang H. B.; Huang X. R.; Li R. X.; Lim T. C.; Ding W. P. "Sound quality prediction of vehicle interior noise using deep belief networks" - Applied Acoustics (2016)
32. Jeong J. -E.; Yang I.-H.; Bin Abu A.; Cha K.-J.; Oh J.-E. "Development of a new sound quality metric for evaluation of the interior noise in a passenger car using the logarithmic Mahala Nobis distance" - Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering (2013)
33. Chen S.; Wang D.; Liang J.; McDonnell M. "Sound quality analysis and prediction of vehicle interior noise based on grey system theory" - Fluctuation and Noise Letters (2012)
34. Wang Y.; Shen G.; Guo H. "Modelling for sound annoyance evaluation of vehicle noise based on neural network" - Przeglad Elektrotechniczny (2012)
35. Hugo Fastl; Eberhard Zwicker "Psychoacoustics, Facts and Models" – Springer (2006)
36. Pedrielli F.; Carletti E.; "Just noticeable differences of loudness and sharpness for earthmoving machines" - The Journal of the Acoustical Society of America (2008)
37. D. Dal Palù; E. Buiatti; G.E.Puglisi; O.Houix; P. Susini; C. De Giorgi; A. Astolfi "The use of semantic differential scales in listening tests: A comparison between context and laboratory test conditions for the rolling sounds of office chairs" - Applied Acoustics (2017)
38. German patent DE3709397C2 – (1987)
39. M. Favaretto; A. Astolfi; L. Shtrepi "Sound Quality evaluation of the interior noise of a tractor HVAC system based on prediction model" (2023)
40. Leo L. Beranek; Warren E. Blazier; J. Jacek Figwer "Preferred Noise Criterion (PNC) Curves and Their Application to Rooms"- The Journal of the Acoustical Society of America 50, 1223 (1971)
41. ANSI/ASA S12.2-2008, American national standard – Criteria for evaluating room noise