

COVER PAGE

Document downloaded by @DAEL

Thu Nov 26 21:43:10 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Sound Quality Assessment of Internal Automotive Road Noise Using Sensory Science, F. Bergeron, C. Astruc, A. Berry and P. Masson, Acta Acustica **vol. 96** (Number 3), 2010, pp. 580-588

DOI
https://doi.org/10.3813/AAA.918308

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 96 (2010) 580 – 588

Technical and Applied Papers
DOI 10.3813/AAA.918308

Sound Quality Assessment of Internal Automotive Road Noise Using Sensory Science

F. Bergeron[1], C. Astruc[2], A. Berry[1], P. Masson[1]
[1] GAUS, Mechanical Engineering Department, Université de Sherbrooke, Sherbrooke QC, J1K 2R1 Canada. Patrice.Masson@USherbrooke.ca
[2] Renault, TCR AVA 1 65, 1 Ave. du Golf, 78288 Guyancourt Cedex, France

Summary
This paper presents a technique to describe the perception of internal automotive road noise. As the acoustic sources on a car reach lower sound levels, special attention is granted to acoustical comfort. While the knowledge of mechanisms behind road noise is growing, its perception by the driver remains relatively unexplored. In this study, a sensory science technique – the classical sensory profile – is applied to obtain a description of the internal road noise by quantitative perceptual criteria. Seven consumer cars (from Renault, Fiat, Peugeot, and Toyota) were driven under four well-defined conditions varying in speed and road surface. Twenty-one typical road noise recordings were presented to a panel of five experts in acoustics. Twelve listening sessions of three hours, individually and in group, led to a list of 15 relevant perceptual descriptors, such as "rough aspect". The performance of the listeners was tested statistically: their repeatability, their ability to discriminate the sounds and the inter-subject agreement. Sufficient agreement was observed across the listeners for 10 of the 15 descriptors, which would constitute the perceptual basis for internal automotive road noise. Through multiple linear regression, strong correlations were found between the associated perceptual dimensions and psychoacoustic properties of the sound samples. The findings from this study are summarized in a sensory grid to assist the test drivers in the evaluation of road noise sound character, and a predictive tool for sound quality evaluation based on correlations between perceptual dimensions and calculated sound metrics.

PACS no. 43.66.Lj, 43.50.Lj

1. Introduction

Sound quality in the automotive industry has attracted considerable interest over the last decades. The impact of a car's acoustics – whether from the engine, road or other components – is known to influence the perceived overall quality from a customer's point of view. Along with the advent of electric cars, internal noise due to engine and powertrain has decreased drastically, accentuating the perception of other noise sources such as road-tyre interaction. It is now common knowledge that customers pay great attention to sound quality even when the sound is only a side effect of the product's operation [1] and that the overall interior sound pressure level is no longer sufficient to assess appreciation [2]. Noise sources must therefore be identified and characterized independently from a perceptive point of view. The prediction and control of the perceived internal road noise is thus a crucial step for manufacturers in order to reach a marketing value which meets the customer's expectations.

Generation of external road noise has been explored by U. Sandberg [3] and divided into low and high frequency ranges. It has been proposed that the low-frequency portion of road noise is due to the impact of tyre thread elements on road chippings, or vice-versa. On the other hand, the high-frequency mechanism is believed to be some air-pumping or air-resonant mechanism perhaps excited by stick-slip motions in the tyre-road interface. To summarize the extreme complexity of the physics of road noise, Sandberg even states that the simplicity and apprehensiveness of presented relations are in proportion to either the limitations of data, or the author's ignorance [3]. Still in this paper, profilometry is suggested as a method to characterize road surfaces regarding the generated road noise. Typically, a laser-beam profilometer is used to record the road's longitudinal profile, whose spectral analysis leads to normalized "texture levels" in four wavelength (λ) bands: microtexture ($\lambda < 0.5$ mm), macrotexture ($0.5\,\text{mm} \leq \lambda < 50\,\text{mm}$), megatexture ($50\,\text{mm} \leq \lambda < 500\,\text{mm}$) and unevenness ($\lambda \geq 500$ mm). Relationships linking spectral components of road noise to the profile texture have been established and are used as an ISO standard for the characterization of pavement texture regarding vehicle noise [4]. In another study, three mechanisms behind the generation of road noise inside the passenger cavity have been proposed by Lalor and Priebsch [5]: direct transmission from the outside through small holes in the cabin walls, mass

Received 18 June 2009,
accepted 28 February 2010.

© S. Hirzel Verlag · EAA

law transmission through cabin walls (excluding their resonant response) and acoustic radiation by the vibrating walls. These three mechanisms would explain the transmission of vibrational and acoustical energy (from the road excitation) to the passenger cavity. The same work presents numerical predictions of interior noise spectral levels that were achieved using finite element and boundary element method, statistical energy analysis, hybrid methods, ray tracing method, and band-averaged transfer function method. While generation and transmission of road noise have been thoroughly investigated, the question of human perception of internal road noise was not addressed as such.

Various methodologies for evaluating sound quality in the automotive industry were proposed by Miskiewicz and Letowski [1]. Two types of criteria for the assessment of sounds were distinguished: sound quality (degree of satisfaction) and sound character (quantitative judgment). Sound character judgment does not imply anything about the listener's preferences or satisfactions, but relies on their perception of a certain aspect (e.g. roughness of an engine noise). Sound quality and character are to be evaluated by either global or parametric assessment. For example, parametric assessment of sound character makes it possible to examine the listeners' particular perception through a set of continuous scales called "perceptual dimensions", whose associated descriptive terms are called "descriptors". For the above-mentioned example, the descriptor would be "engine sound roughness" and its perceived magnitude would be evaluated by a listener, on a scale representing this specific perceptual dimension. In order to fully characterize the perception of a complex sound, one needs several perceptual dimensions, each of which representing the various "aspects" of the sound character. The object of the present study could therefore be defined as a sound character parametric assessment for internal automotive road noise.

Several studies have linked the satisfaction of a set of customers to psychoacoustical metrics computed on recorded sounds. Schiffbänker *et al.* [6] developed a methodology based on paired-comparison tests, principal component analysis and multiple regression stepwise analysis, applied to the prediction and reduction of the unpleasant impression in engine noise. Otto and Wakefield [7] studied the preference of automotive starter sounds using paired comparisons. Dedene *et al.* [8] also used paired comparison in addition to multiple linear regression to predict objectively the perception and appreciation of exhaust sounds. In these cases, the choice of the evaluation methods was justified by the type of listeners (not sound quality experts) and the goal of their work, assessing customer satisfaction through sound quality analysis.

Excluding the preference criterion, several studies addressed sound character in the automotive industry. Wakita *et al.* [9] proposed a one-parameter sound character assessment of the engine rumble noise using a panel of twelve listeners and synthesized engine sounds. Barthod *et al.* [10] studied the perception of noise known as "rattle" in gearboxes, using the dissemblance test technique. The authors compared a list of known significant factors to the variance portrayed by the parametric evaluation of the rattling sound aspect. Dedene *et al.* [8] also examined the correlation between sound perception of exhaust sound with sound quality metrics, again proposing an established set of parameters (sportiveness and pleasantness). Lee and Park [11] linked subjective evaluation of "booming" and "rumbling" perceptions to sound "indices" (objective evaluations) using artificial neural network. Finally, Hashimoto [12] applied a correlation technique to describe the perception of booming and impulsion of engine noise in terms of usual psychoacoustical parameters. From this, it is clear that most of the studies related to sound character in automobiles propose in advance a known set of sound descriptors (e.g. rumble, rattle, sportiveness, boom…). The listeners are then asked to evaluate the magnitude of the corresponding perceptions using linear scales ("perceptual dimensions"). In the above cases, the underlying sound descriptors were precisely known to the listeners, such that they anticipate what to seek for. In the case of road noise, however, such descriptors are not widely known or accepted as universal. Primarily due to road noise's intricacy, these descriptors, whose relevance has never been confirmed, are usually noted subjectively and intuitively. This should explain why very little has been accomplished regarding sound quality or character of the perceived road noise: its complexity makes its evaluation challenging.

This study presents the results from an approach implemented to derive a list of descriptors for the characterization of internal road noise. Acoustical metrics are then correlated to the associated perceptual dimensions to propose a prediction of the perception as assessed by a human being. It should be noted that the presented methodology leads to a set of descriptors whose relevance has been thoroughly tested on a small number of sound experts. In this view, the results relate to this specific group and would require extensive validation if they were to be used by other evaluators or in another context (i.e. inducing cultural, experience, or language biases). Therefore, although the derived road noise descriptors and models constitute the main results of this study, emphasis is brought to the methodology and analysis of results, which can be applied to a variety of contexts in order to assess the perception of automotive noise.

2. Sensory analysis protocol

Sensory analysis methods have been extensively developed in the mid-1900's by the food industry, and have also been applied to the cosmetics, textile, pharmaceutical and transportation fields. The classical sensory profile is one of the descriptive methods of sensory science which allows describing all of the sensorial properties of consumer products [13, 14, 15]. This method has been described and successfully implemented in acoustics, for example on air-conditioning noise by Siekiersky *et al.* [16, 17]. The de-

scription is based on the perception of a small group of trained listeners. The final goal of a classical sensory profile is to obtain an evaluation grid and a sensory positioning of the tested products in order to compare and describe them. In the case of this study, five acoustics experts participated in twelve weekly listening sessions, each of which lasting approximately three hours. The choice of experts in acoustics (as opposed to non-expert customers) is motivated by their greater elocution regarding their acoustical perception and by the availability of such experts. Also, the final output being a tool for sound character evaluation, and not for customer appreciation assessment, it makes sense for a car manufacturer to choose technicians and engineers in acoustics as the listeners participating in the sensory profile.

2.1. Measurement set-up

A collection of sound samples was to be constructed from real-life situations where road noise prevailed. One critical step was to ensure that the presented sound samples were exhaustive and representative of the road noise usually heard by consumers. For this purpose, twenty-one different sounds were recorded in seven consumer vehicles (Renault Espace, Renault Modus, Renault Clio, Renault Mégane, Fiat Panda, Peugeot 207, and Toyota Auris) according to four driving conditions. Following the recommendations from ISO standards [4], the driving conditions were specified in order to enhance road noise variations by combining two parameters known to influence road noise: pavement microtexture and macrotexture. As summarized in Table I, four road surfaces (rough, protruding, smooth, highway) were associated with four driving speeds (60, 70, 80, 100 km/h), combining low and high levels of pavement micro- and macrotexture. The difference in speed between the driving conditions induced a variety of excitations caused by very large profile wavelengths (megatexture and unevenness). Through their different features (type of suspension, rim material, presence of a sunroof…), the seven vehicles were also selected in order to generate a variety of road noises.

The sound pressure signals were sampled at 51.2 kHz and digitally recorded using the Head HMS binaural measurement system in the passenger seat. This equipment is of great importance as it emulates the transmission characteristics comparable to human external hearing system [2]. Special care was taken to ensure that the road noise was predominant. Regarding engine noise, in all cases, the transmission was set to neutral. In addition, in highway conditions, where aerodynamic noise is an important contributor to interior noise, we ensured that no wind gusts (high-speed and unstationary flow) were included in the final recordings. From this sound bank, a limited number of 21 samples were selected for the listening tests. These were chosen for their variety and their signal cleanness. The sounds were then leveled to a common psychoacoustic loudness value [18], as this variable is known to be the main factor influencing sound perception [6]. Eliminating loudness as a variable brings conclusions based on

Table I. The four driving conditions are combinations of low and high levels of road surface textures.

	Low macrotexture	High macrotexture
Low microtexture	Smooth track, 80 km/h	"Protruding" surface, 70 km/h
High microtexture	Highway, 100 km/h	Rough track, 60 km/h

Table II. List of the twelve listening sessions, grouped in three stages, constituting the protocol of the classical sensory profile. Each session took place weekly and lasted approximately three hours. t+a: terms and anchors. Stage 1: Generation of descriptive terms. Stage 2: Training in magnitude rating. Stage 3: Final evaluation

Session	Sensory profile session content		Type
Session 1	Stage 1	– Generation of terms	Individual
Session 2		– Qualitative sorting	In-group
Session 3		– In-group discussion	In-group
Session 4		– Quantitative sorting	Individual
Session 5		– Quantitative sorting	Individual
Session 6		– Definitions of t+a	In-group
Session 7		– Validation of t+a	In-group
Session 8	Stage 2		Individual
Session 9			Individual
Session 10	Stage 3		Individual
Session 11			Individual
Session 12			Individual
Presentation of results			

sound aspects rather than levels, which answers more adequately the underlying issue of sound character parametric assessment. The sessions were conducted in a room that allows simultaneous listening by the five experts, which facilitated the discussion between them. For each listener, a pair of high-quality STAX SR303 open headphones was made available for the medium and high frequency reproduction, the lower part of the spectrum being emitted by a Haeliatus HA-224 subwoofer. This ensured adequate reproduction over the entire frequency range. It should be specified that the listeners had visual access to a screen showing only 3-digit sound codes that were modified for every listening session.

2.2. Generation of descriptive terms

The classical sensory profile was conducted under three main stages, divided into 8 individual and 4 in-group listening sessions (shown in Table II), the first stage being the generation of descriptive terms. The listeners were asked to individually write on a sheet of paper the lexicon they would use to describe and differentiate the sound samples that were presented through binaural playback. In order to accommodate the majority of listeners, each 5-second sample was looped three times, followed by a 10-second silence. This 25-second segment was repeated (6 times on average) until the listeners agreed that their individ-

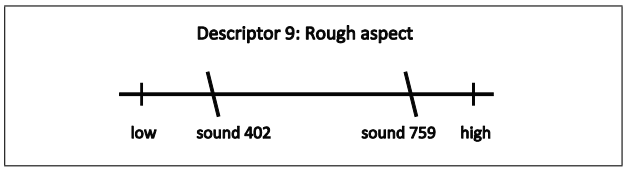


Figure 1. Example of a linear scale used by the listeners to evaluate the magnitude of a perception between "low" and "high" limits. Actual scales used were longer (15 cm).

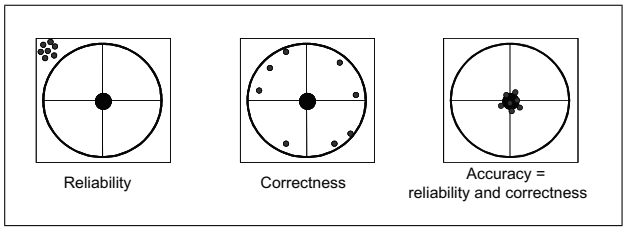


Figure 2. Illustration of the reliability, correctness and accuracy of a measuring device.

ual list of descriptive terms was exhaustive for the presented sound. This process was repeated for the 21 road noise samples. Qualitative sorting was then accomplished from in-group listening and discussion (c.f. Table II: session 2) and had for objective to eliminate hedonic terms (preferences), synonyms and antonyms. The following in-group discussion concerning the definition of the terms (session 3) allowed the listeners to agree on terms which first carried a different meaning among them. Remaining non-obvious synonyms were identified and eliminated by the panel during this discussion. The reduced list of terms was used in sessions 4 and 5, where listeners were asked to evaluate the perceived magnitude of each descriptor, on each sound, on discrete 6-step scales ranging from 0 (no intensity) to 5 (strong intensity). The objective of the subsequent quantitative sorting was to reduce this list by identifying the less used terms and the correlated terms [19]. Two statistical indicators were used for processing the quantitative data which consisted in the 0-5 scores from each listener, on every sound. First, the geometric mean, reflecting the usage of a descriptor, helped find terms that were used only for a minority of sound samples. Then, principal component analysis (PCA) was achieved in order to reveal high correlations between descriptors. This type of analysis allows visualization and interpretation of data by associating the greatest amount of observed variability to a reduced number of variables [15]. PCA permitted the quantitative validation of the assumption that all descriptors were used by every panelist and that strong oppositions or redundancies did not occur. Lastly, in sessions 6 and 7, two "anchor sounds" (low and high sound references) were chosen for each descriptor, to insure a consistent output on linear scales [20], eventually used in the next stage. At this point, a list of "potential candidates" for the descriptors of internal road noise was drawn. These descriptors were said to be relevant, precise, discriminating, independent, exhaustive and neutral (without preferences).

2.3. Training in magnitude rating

The second stage consisted in the training of the experts in magnitude rating. For each descriptor, the listeners evaluated the magnitude of the associated perception using linear continuous scales, marking the perceived magnitude of each sound (see Figure 1). For this purpose, a sheet containing the entire set of descriptor scales was provided to each listener. After the playback of each 25-second sound segment, the listeners were asked to draw a crossing line representing their perceived magnitude on every descriptor's associated scale. The score was measured as the distance in centimeters from the lower end of the 15-cm scale. The advantages of this magnitude estimation method are the absence of a "ceiling effect" and the theoretical infinite resolution [20]. The two sessions (Table II, sessions 8–9) of this training phase brought the experts to use the total length of the scales, thus accentuating their discriminating power. This is the main objective of the training phase, as the actual results are not further processed.

2.4. Final evaluation

The third stage of the classical sensory profile consisted in three sessions of magnitude rating. For each of these sessions, the experts were asked to evaluate the complete set of sounds for all of the sensory descriptors, again on linear continuous scales using the same low and high sound anchors. The two main objectives of this stage are to assess the panel's statistical performance using the relevant descriptors and to collect perceptual data on each sound.

In sensory analysis, since the experts are to be considered as measuring devices, they must be reliable, correct and accurate. As illustrated in Figure 2, reliability is the ability of a subject to produce, in specified conditions, similar responses to similar stimuli over time. Correctness describes a subject for whom the average response is very close to the known, or expected value. An accurate subject possesses the characteristics of both reliability and correctness.

On relevant descriptor scales, the average of the individual measurements of a group determines the true or expected value of this measurement. As the responses usually vary slightly from a subject to another, it is necessary to collect the responses from several trained subjects. Consequently, three successive sessions of evaluation were accomplished to test the performance of the listeners. Their individual repeatability, their ability to discriminate the sounds and their mutual agreement were tested throughout the third stage of the sensory profile. Statistical techniques such as standard deviation, analysis of variance and principal component analysis, applied to the data collected in the final evaluation, led to conclusions regarding the performance of the panel and the relevance of the selected descriptors.

Furthermore, for sessions 8-12, the order of listening was randomized for every session using the latin square technique, thus minimizing the possible interaction between subsequent sounds [21]. At the end of this third and

final stage of the sensory profile, the collected data can be used to assess the relevance of each descriptor and the accuracy of each listener.

3. Correlation with objective metrics

The following step aimed at correlating the panel's mean evaluated magnitude of each perceptual dimension (quantitative scale linked to a descriptor) to a combination of known objective sound metrics. The analysis should provide a set of metrics that is:

- Specific to each perceptual dimension.
- Representative of the average assessment of the experts and contained in the experts' standard deviation.
- Linked to perceptual phenomena documented in (psycho-) acoustics literature.

Additionally, it should be specified that the aim of the analysis was not to compare the validity of existing psychoacoustic quantities, nor to find the combination of objective metrics which provided the strongest correlation with the experts' mean evaluation, regardless of the number of implied variables. The proposed set of metrics should be as simple as possible in order to ease its interpretation and express a direct link between each perceptual dimension and its physical cause. The proposed calculations included the usual (psycho-)acoustic metrics [9, 22, 18, 23]:

- Octave, 1/2 octave, 1/3 octave and arbitrary band averaged sound pressure levels (SPL),
- A, B, C and D weighted sound pressure levels,
- Zwicker loudness on 24 Bark bands,
- Roughness and sharpness levels, fluctuation strength, modulation energy,
- Impulsiveness, tone-to-noise ratio (TNR),
- Spectral centroid, absolute deviations, kurtosis, skewness.

In order to achieve correlations between the subjective perceptual dimensions and objective metrics, a multiple linear regression analysis was conducted. For a given perceptual dimension, we propose a set of n metrics computed on the 21 sounds. As each perceptual dimension is independent, separate analyses were led, such that the selection and number of calculations n varied among the perceptual dimensions. Not all of the above-mentioned metrics were included in every analysis; only those that primarily showed relevance to the analyzed descriptor. The values of each metric x is calculated on the 21 sounds, forming a matrix $(x_{1,1} \dots x_{21,n})$ which, when multiplied by coefficient vector $\{a_0 \dots a_n\}$, yields predicted result vector $\{\bar{y}_1 \dots \bar{y}_{21}\}$,

$$\begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \cdots \\ \bar{y}_{21} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,n} \\ 1 & x_{2,1} & \cdots & x_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{21,1} & \cdots & x_{21,n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_n \end{pmatrix}.$$

The results of the third stage (final evaluation) of the sensory profile were averaged between the listeners. These 21 quantities were considered to be the true or expected values $\{y_1 \dots y_{21}\}$ of the perceived magnitudes for this perceptual dimension. The optimal set of metric coefficients $\{a_0 \dots a_{n1}\}$ minimizes the sum of the squared residuals calculated as a difference between the prediction $\bar{y}_i$ and the expected value y_i,

$$\sum_{i=1}^{21} \left(y_i - \bar{y}_i\right)^2 = \sum_{i=1}^{21} \left(y_i - (a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \cdots + a_n x_{i,n})\right)^2.$$

For this purpose, a least-square algorithm was used to solve the above matrix system for optimal coefficients $\{a_0 \dots a_{n1}\}$. The quality of the regression was illustrated by the corresponding coefficient of determination R^2 computed between the predicted and expected values. The outcome of this analysis is a set of coefficients that, matched with known metrics, yields predicted perceptual magnitudes with a high degree of statistical confidence.

Many of the proposed quantities required specific frequency limits to constrain the processing in particular spectral areas. For example, the sound pressure level in an arbitrary frequency band can be framed between a "lower" and a "higher" limit frequencies. On analyses where only one metric is proposed, these optimal values can be found by simple graphical optimization. By iterating the regression analysis and mapping the coefficients of determination (R^2) as a function of the two limit frequencies, one can visualize the optimal combination of lower and higher limit frequencies.

4. Results of the sensory analysis

4.1. Generated lexicon and performance of the panel

The objective of the first stage of the classical sensory profile was to generate a list of terms that would be used to describe and evaluate road noise. During the exhaustive generation of terms (Table II, session 1), the five experts listed 183 terms describing the 21 sounds, varying from 17 to 60 terms by expert. The number of terms from two experts was significantly higher, as they used many analogies and synonyms to describe their perception. These 183 terms were then clustered, by semantic meaning, into 37 groups. From this clustering and the following in-group discussions, the experts reduced the number of results to 27 relevant descriptors which would be tested on 6-step discrete scales (sessions 4–5). From hypothesis testing, the main results of the quantitative sorting showed that, under a statistical confidence level of 95%,

- For 11 of the 27 terms, the geometric mean on the 0-5 scores was lower than 50%, revealing either an infrequent use of the term by the listeners or a low occurrence of the corresponding perceptions in the sound bank. This did not automatically lead to eliminating these descriptors, since the performance of the panel was to be increased through training.
- 20 terms allowed differentiating sounds in a statistically significant way, such that seven were eliminated.

Table III. Descriptors included in the final evaluation of the sensory profile. Following the final evaluation, descriptors in bold were considered the most relevant to the evaluation of the perception of road noise (see sections 4.1.1., 4.1.2 and 4.1.3).

	Original French term	Closest English translation
D1	Aspect typé	Characteristic aspect
D2	Intensité des très basses fréquences	Very low frequency content
D3	Intensité des basses fréquences	Low frequency content
D4	Intensité des moyennes fréquences	Medium frequency content
D5	Intensité des hautes fréquences	High frequency content
D6	Bruit de souffle	Blow noise
D7	Fluctuations en hautes fréquences	High frequency fluctuations
D8	Aspect sourd	Muffled aspect
D9	Aspect rugueux	Rough aspect
D10	Aspect feutré	Felty aspect
D11	Aspect grondant	Rumbling aspect
D12	Aspect hululant	Hooting aspect
D13	Aspect tambourinant	Drumming aspect
D14	Aspect chuintant	Hissing aspect
D15	Aspect résonant	Resonating aspect

- Two terms were obvious antonyms according to the definitions of the experts and to the significant negative correlation shown by PCA. Two other terms were found to be positively correlated. Therefore two terms were hereby eliminated.

Three of the 18 remaining descriptors were eliminated through the following in-group discussions (sessions 6 and 7), because no consensus could be achieved regarding common definitions and (high or low) anchor sounds. Table III presents the 15 descriptors that were included in the final evaluation. As the study was conducted in French, the original terms are included in the table since semantic precision may be lost through translation to English.

Following the training in magnitude rating (sessions 8–9), results from the three sessions of final evaluation were collected and processed to measure the repeatability, discriminating power and agreement of the group.

4.1.1. Individual repeatability

In sensory analysis, the role of an evaluator is to estimate sound character in a most reliable way, such that the measured response does not change with time. This dispersion between responses is illustrated by the standard deviation computed on subsequent assessments. As shown in Table IV, the standard deviations between the three responses of each listener were significantly low. On average, expert E4 was the most reliable, followed by E1, E2 and E5, and lastly, E3. In the third evaluation session, the listeners were presented 22 sounds, one of which being presented twice. The experts were not aware of this repetition; they were merely informed that a new sound had been added. A Pearson's correlation test [24] (hypothesis testing based on correlation coefficients) within the 15 evaluations on the two identical sounds and for each expert showed that the correlations were significant, attesting intra-session repeatability. In conclusion, except for expert E3, the listeners were thus generally reliable during the same session and throughout the procedure (intra-session and inter-session repeatability).

4.1.2. Discriminating power

The discriminating power of a sensory descriptor refers to the ability of a subject to differentiate the sounds with respect to its corresponding perceptual scale. For this purpose, analysis of variance (ANOVA) was used. This technique is a collection of statistical procedures, commonly used in design of experiments, in which the observed variance is attributed to various explanatory variables [21]. In sensory science, it is mainly used to describe the level at which the observed variance in the perceptual data can be attributed to a subject (listener), a product (sound), or a combination of both. In our case, one-way product analysis of variance was conducted to verify the discriminating power of the sensory descriptors. Results are shown in Table V. Except for descriptor D10, experts E1, E2, E4 and E5 differentiated the sounds for all of the descriptors. The low reliability of expert E3 is translated in a lack of discriminating power, validating the relevance of 14 descriptors for the four other experts.

4.1.3. Inter-subject agreement

The last statistical validation of the 15 descriptors measured the mutual agreement between the evaluated perception between the listeners. A factorial discriminant analysis [24] applied to the mean assessments on the 15 descriptors led to the conclusion that the driving condition had a significant effect on the evaluation. Regarding this, two-way (product by subject) ANOVAs, including interactions, were thus conducted separately for each driving condition. The results were as follows, for a statistical confidence interval of 95%:

- For all descriptors, a significant subject effect indicated that the experts used either the low or the high part of the scale. Despite several training sessions, this scale effect was shown by significant subject effects.
- A significant product effect showed that the group perceived the 21 sounds as different, thus confirming the initial hypothesis stating that the sounds should present a notable diversity.
- On several descriptors, non-significant product-by-subject interaction effects indicated that the positioning of sounds in magnitude was identical among the listeners. However, disagreement within the experts occurred for the descriptors which showed high product-by-subject interaction effects. The 5 descriptors which showed the weakest agreement among the experts were eliminated.

From these results, it was clear that 10 of the 15 descriptors showed sufficient relevance for the selected panel of listeners.

Table IV. Repeatability of the five listeners' evaluations as a measurement of their individual reliability. The numbers represent the average standard deviation between the three final evaluations computed over the set of sounds. Scores are recorded in centimeters on continuous 15-cm scales. In black, standard deviation $\sigma = 1.5$ is considered acceptable; in light grey $\sigma = 3$ is considered unacceptable and in dark grey, $1.5 < \sigma < 3$.

Listener	Average	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Expert 1	1.7	2.1	2.0	1.5	1.7	1.4	1.9	1.6	1.2	1.6	1.9	2.0	2.0	1.7	1.7	1.7
Expert 2	1.7	1.3	1.9	1.6	1.7	1.3	2.3	1.5	1.5	1.8	1.1	2.6	1.5	2.4	1.8	1.7
Expert 3	2.1	2.4	1.6	2.2	3.0	2.0	2.0	2.7	2.0	2.1	2.0	2.5	2.1	1.3	2.1	2.0
Expert 4	1.4	1.7	1.3	1.3	2.2	1.3	1.3	0.8	1.3	1.4	1.3	1.8	1.0	1.1	1.6	2.1
Expert 5	1.7	2.6	1.1	1.6	1.3	1.8	1.6	1.9	1.9	1.5	2.7	1.4	1.4	1.3	1.7	2.3
Average	1.8	2.0	1.6	1.6	2.0	1.6	1.8	1.7	1.6	1.7	1.8	2.1	1.6	1.6	1.8	2.0

4.2. Sensorial positioning of sounds

This section presents results summarizing the sensory information on a sensory map where the sounds are positioned according to their perceptual magnitudes. We conducted one PCA per driving condition, using the average magnitude outputs from listeners E1, E2, E4 and E5 as the true perceptual magnitudes. The example of the "protruding surface driving condition" is illustrated by Figure 3. The two axes F1 and F2 (the two principal components) represent the two linear combinations of perceptual dimensions which contribute to most of the observed variance in the data. Each point represents the projection of a sound's perceptual magnitude on the principal plane. In this case, 81.11% of the variance between the magnitude ratings results from these two principal components (47.16% for the first one and 33.95% for the second). From the graph on the left, one can easily deduce that since the five automobiles (sounds) are well spread over the four quadrants of the plane, there were notable differences in the associated perception (they sound radically different). This confirms again the underlying hypothesis regarding the perceptual diversity of the recordings. In a case where all of the points would be grouped in the same quadrant, one could state that the sounds or the perceptual dimensions are not adequately discriminant to draw relevant conclusions from PCA.

The circular graph on the right shows the positioning of each descriptor (vector) plotted against principal components F1 and F2. If the length of such a vector is 1, then this descriptor can be completely quantified by a linear combination of F1 and F2. Vectors whose lengths are close to unity and which are collinear represent strong (negative or positive) correlation. In this example, D11 ("rumbling aspect") and D2 ("very low frequency content") showed this behavior, as the information from one descriptor was redundant to the other. However, since separate analyses were led for each driving condition, this may not mean that D11 and D2 are always redundant or synonyms.

A more exhaustive interpretation of the various PCAs obtained in this road noise study can be found in [25]. The sensorial positioning of the sounds confirmed that in each condition, considerable differences between the sounds occurred, validating the variety of recorded noises. Moreover, the identification of the dominating sensory descriptors for one condition will be beneficial to the characterization of the vehicles regarding some specific quality target.

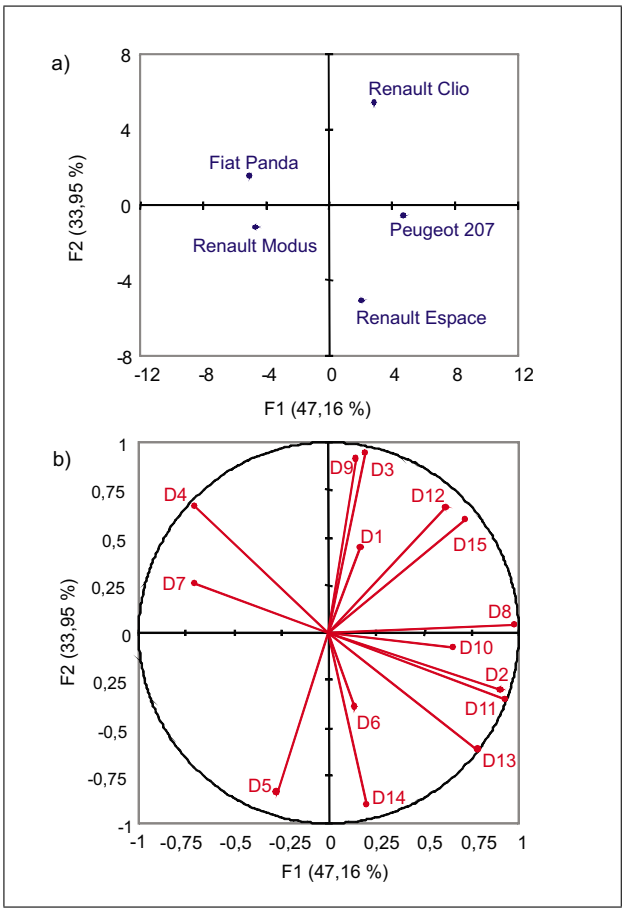


Figure 3. Results from the Principal Component Analysis on the protruding surface driving condition. a) sensorial positioning of the five sound samples on the principal plane. Values are projected magnitude scores in centimeters, on a 15-cm scale. b) positioning of the 15 descriptors on the principal plane. Values are fractions of unity.

Through the results regarding individual repeatability and discriminating power it can be stated that four of the listeners showed sufficient performance to be considered as accurate measuring devices. The lack of inter-subject agreement on five sensorial descriptors reduced the list to 10.

Table V. Results of the one-way product ANOVA on the listeners' evaluations. Stars imply a significant effect for an interval of confidence of 95% ($p = 0.05$); *: < 0.05, **: < 0.01, *** < 0.001; dark grey numbers mean a significant effect for an interval of confidence of 90% ($p = 0.10$) and light grey numbers show an insignificant effect ($p > 0.10$).

Listener	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
Expert 1	**	***	**	*	***	**	**	***	***	0.01	**	***	***	***	**
Expert 2	*	***	***	***	***	***	***	***	***	0.68	***	***	**	***	***
Expert 3	***	***	**	0.15	***	0.15	0.17	***	***	0.15	***	***	***	***	0.45
Expert 4	**	***	***	*	***	***	***	***	***	***	***	***	***	***	**
Expert 5	*	***	***	***	***	***	**	***	***	*	***	***	***	***	***

5. Results of the regression analysis

This section presents the statistical results to the regression analysis linking sensory dimensions to (psycho-)acoustic metrics. Only the magnitudes collected in the evaluation stage, from the four accurate experts and the 10 relevant dimensions, were considered. For all perceptual dimensions, only one or two objective metrics were included in the multiple regression analysis. Stepwise reduction algorithms are then unnecessary, and links can be found directly from signal processing. An example of the regression results for dimension D5 ("high frequency content") is shown in Figure 4.

In this example, the predicted data, using only one metric (sound pressure level averaged over the [500, 4000] Hz band), always stand within one standard deviation from the panel's average assessment. Table VI shows the proposed objective metrics associated to each perceptual dimension. As seen in Table VI, from the 10 selected perceptual dimensions, eight were quantified with coefficients of determination R^2 greater than 0.8. In these cases, it can be concluded that the calculated results account for at least 80% of the total variations observed. In all cases, p-values [24] lower than 0.05 confirmed that the correlation was conducted under high statistical confidence. Among the other two calculations, one (D12, "hooting aspect") depended upon systematic user input from an analyst. This situation occurred because calculating the tone-to-noise ratio required manual detection from a human being. The remaining dimension (D4, "medium frequency content") showed strong data polarization, which translated into an "yes/no" behavior and raised problems in regression. Using the described graphical optimization technique, the multiple regression analysis led to an optimal set of objective metrics which were correlated to the average perceived magnitudes on 8 of the 10 relevant perceptual dimensions.

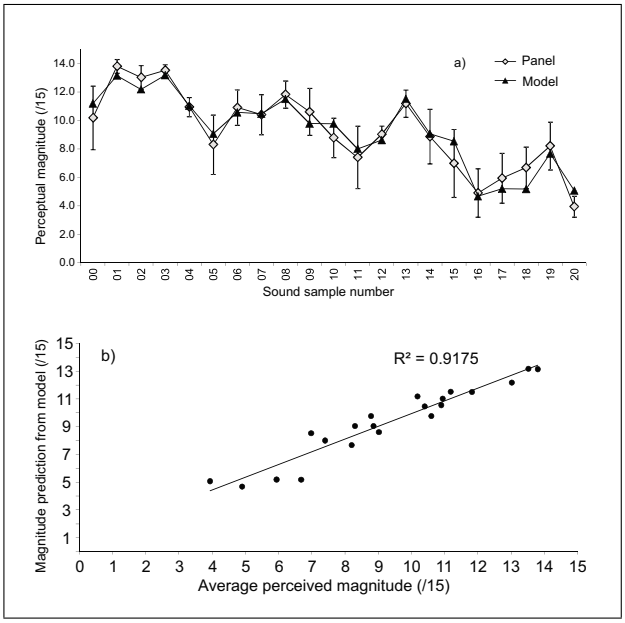


Figure 4. Example of the regression analysis on dimension D5 (high frequency content), using one metric (SPL averaged over the [500, 4000] Hz band). a) comparison of mean assessments and predicted results on the 21 sounds, error bar : standard deviation between the listeners. b) same data plotted in increasing order of average rating.

Table VI. Coefficients of determination associated to the 10 proposed predictive calculations.

	Proposed calculation	R^2
D2	[15, 60] Hz dB(A) SPL	0.89
D3	[60, 250] Hz dB(A) SPL	0.87
D4	[250, 500] Hz dB(A) SPL	0.44
D5	[500, 4000] Hz dB(A) SPL	0.92
D8	[15, 600] Hz Spectral centroid	0.86
D9	[90, 200] Hz Roughness	0.83
D11	[40, 180] Hz Roughness	0.80
D12	TNR (requires human input of frequency ranges)	N/A
D13	Combination of [15, 60] Hz dB(A) SPL and [15, 60] Hz impulsiveness	0.96
D14	[800, 4000] Hz dB(A) SPL	0.93

6. Conclusion

This study has demonstrated a successful application of sensorial techniques to achieve a list of 10 perceptual dimensions describing the sound character of interior automotive road noise. The results have shown that the listeners' evaluations are repeatable, such that they assign the same magnitude ratings from a listening session to another. Furthermore, they differentiate the 21 road noises, since the sensory descriptors have a significant discriminating power. In spite of the differences observed between the experts, their level of precision was considered suffi-

cient as an output of the sensory profile, for 4 experts and 10 perceptual dimensions. For eight of these, multiple linear regression allowed to identify a numerical model of the experts' perception (with $R^2 \geq 0.80$).

This study helped creating a new on-track evaluation grid for the assessment of internal automotive road noise. In addition, a prediction of the expected perceived sound character parameters can now be accomplished through an associated software. In the near future, the sensory grid will require validation in real-life situations, as the sensory profile was based on room listening. The experts should evaluate the same vehicles in the same driving conditions as the sound recordings, and their assessments should be compared to those of this study.

In conclusion, coupling sensory methods with regression analysis allowed to significantly improve the understanding of the perception of internal road noise. From the associated software, a prediction of the human perception can be integrated to sound quality oriented design. Such a sensory technique can be extended to other acoustic components of a vehicle, or even to other human senses participating in global automotive quality assessment.

Acknowledgement

We would like to acknowledge the help of five experts from Renault for their implication in this work: Terry Alphonsine, Sonia Debant, Yannick Denoual, Kevin Dos-Santos and Laurent Galtier. Partial financial support from the AUTO21 network is also acknowledged.

References

[1] A. Miskiewicz, T. Letowski: Psychoacoustics in the automotive industry. Acta Acustica united with Acustica **85** (1999) 646–649.

[2] K. Genuit: The sound quality of vehicle interior noise: a challenge for the NVH-engineers. Int. J. Vehicle Noise and Vibration **1** (2004) 158–168.

[3] U. Sandberg: Road traffic noise - the influence of road surface and its characterization. Applied Acoustics **21** (1987) 97–118.

[4] International Organization for Standardization. 13473: Characterization of pavement texture by use of surface profiles. 1997–2008.

[5] N. Lalor, H.-H. Priebsch: The prediction of low- and mid-frequency internal road vehicle noise: a literature survey. Proceedings of the Institution of Mechanical Engineers - Part D: Journal of Automobile Engineering **221** (2007) 245–269.

[6] H. Schiffbänker, F. K. Brandl, G. E. Thien: Development and application of an evaluation technique to assess the subjective character of engine noise. SAE Technical Paper Series #911081 - Proceedings of the 1991 Noise and Vibration Conference, Traverse City MI, May 13–16, 1991.

[7] N. C. Otto, G. H. Wakefield: A subjective evaluation and analysis of automotive starter sounds. Noise Control Engineering Journal **41** (1993) 377–382.

[8] L. Dedene, R. Valgaeren, M. Van Overmeire, P. Guillaume: Correlation of subjective sound perception of exhaust systems with sound quality metrics. SPIE proceedings series. MAC: International modal analysis conference No16, Santa Barbara CA, February 2, 1998, 1007–1013.

[9] T. Wakita, Y. Kozawa, K. Samada: Objective rating of rumble in vehicle passenger compartment during acceleration. SAE Papers #891155, May 1989.

[10] M. Barthod, B. Hayne, J.-L. Tebec: Auditory perception of noise known as "rattle" in gearboxes. Acta Acustica united with Acustica **91** (2005) 180–191.

[11] S.-K. Lee, D.-C. Park: Sound quality evaluation in passenger cars based on a two-dimensional booming and rumbling sound index. Noise Control Engineering Journal **55** (2007) 457–465.

[12] T. Hashimoto: Sound quality approach on vehicle interior and exterior noise. Quantification of frequency related attributes and impulsiveness. Journal of the Acoustical Society of Japan [E] **21** (2000) #6.

[13] International Organization for Standardization. 13299: Sensory analysis - Methodology - General guidance for establishing a sensory profile. 2003.

[14] H. Stone, J. L. Sidel: Sensory evaluation practices. Academic Press, San Diego, 1993.

[15] M. C. Meilgaard, G. V. Civille, B. T. Carr: Sensory evaluation techniques. CRC Press, Boca Raton, 1999.

[16] G. V. Civille, J. Seltsam: Sensory evaluation methods applied to sound quality. Noise Control Engineering Journal **51** (2003) 262–70.

[17] E. Siekiersky, C. Derquenne, N. Martin: Sensory evaluation of air-conditioning noise: sensory profiles and hedonic tests. Proceedings of the 17th International Congress on Acoustics, Rome, Italy, 2001.

[18] International Organization for Standardization. 532: Acoustics - Method for calculating loudness level. 1975.

[19] International Organization for Standardization. 11035: Sensory analysis - Identification and selection of descriptors for establishing a sensory profile by a multidimensional approach. 1994.

[20] H. F. J. Blauert: Communication acoustics. Springer, 2005. Chapter 6, 139-160.

[21] D. C. Montgomery: Design and analysis of experiments, 6th ed. John Wiley and Sons, Inc., 2005.

[22] H. Fastl: The psychoacoustics of sound-quality evaluation. Acustica united with Acta Acustica **83** (1997) 754–764.

[23] Zwicker, Fastl: Psychoacoustics: Facts and models. Springer, 2000.

[24] M. O'Mahony: Sensory evaluation of food. Statistical methods and procedures. Dekker, Inc., New York, 1986.

[25] C. Astruc, F. Bergeron, L. Galtier: The use of sensory science to understand sound quality concerning road noise. Proceedings of the 5th Symposium of Automobile and Railroad Comfort of the French Automotive Engineers Society, Le Mans, France, November 19-20, 2008.