



Sound quality of side-by-side vehicles: Investigation of multidimensional sensory profiles and loudness equalization in an industrial context

Abdelghani Benghanem^{1,2,*}, Olivier Valentin^{1,2}, Philippe-Aubert Gauthier^{2,3}, and Alain Berry^{1,2}

¹ Groupe d'Acoustique de l'Université de Sherbrooke, 2500 Boul. de l'Université, Sherbrooke, J1K 2R1 Québec, Canada

² Centre for Interdisciplinary Research in Music, Media, and Technology, McGill University, 527 Sherbrooke St. West, Montréal, H3A 1E3 Québec, Canada

³ École des Arts Visuels et Médiatiques, Université du Québec à Montréal, 405, rue Sainte-Catherine Est, Montréal, H2L 2C4 Québec, Canada

Received 3 January 2020, Accepted 15 December 2020

Abstract – The sensory perception of products influences the relationship of potential users or buyers with these products. Sound quality is part of this sensory experience and is critical for products such as sports or utility vehicles as the sound conveys the impression of power or efficiency, among others. Therefore, there is a need to provide tools based on scientific methodology to acoustical engineers designing such vehicles. The motivation of this work was the need to explore new and faster methods for quicker and simpler sound quality evaluation. In this paper, the sound quality of side-by-side utility vehicles is investigated using the rapid sensory profile measurement method, and then by creating virtual participants using bootstrapping methods. Additionally, this study also investigates the effect of loudness equalization of the sound samples used during the listening tests. Results from these studies were used to establish the sensory profiles, desire-to-buy values and desirable sound profiles regarding the tested vehicles. Equalized loudness tests provide a finer sensory profile than those obtained using non-equalized sound samples. Furthermore, statistical analysis results confirm that adding virtual participants to the original data using a bootstrapping approach helps highlighting key information without altering the validity of the results.

Keywords: Sound quality, Sensory profiles, Loudness equalization, Bootstrap, Sound design

1 Introduction

The sounds of consumer products influence the general perception of these products. This is particularly important for vehicle interior sound and therefore for vehicle manufacturers who wish to attach a positive impression to their products and to improve users satisfaction. In this vein, the field of sound quality has been expanding rapidly in the past decades [1]. Sound quality is even more significant for sport or utility vehicles as their sounds should either convey sportiveness or an impression of power or efficiency. In this paper, we are interested in understanding the perceived sound quality of side-by-side vehicles (SSV).

The SSVs are four-wheel drive recreational vehicle of the utility type. They are called side-by-side because they usually have two seats beside each other. Such vehicles are used to work on farms or in forests in rugged terrain, among other things. SSVs are designed to carry a driver and a passenger seated on his right and are equipped with a dump body, located at the rear of the passenger com-

partment. The vehicle can be used to transport loads in its dump box (tools, parts, furniture, materials such as earth, sand, and so on.) and to move trailers. These vehicles are also equipped with a cabin that allows the passenger compartment to be closed to improve user comfort in cold or bad weather [2]. Because they are often used for heavy works, the sensation of power is important for the customers and makers, a sensation that can be communicated through sound, hence the present work. For 95% of usage time, these vehicles are used at constant speed for transportation.

This study is part of a larger applied research project that combines physical and perceptual studies of these vehicles with an industrial partner. The specific context for this work also adds practical constraints, develop and test a rapid, less time-consuming, pragmatic, workflow for sound quality studies in realistic industrial context where quick answers are required in order to adapt the acoustical design in a yearly cycle of design and fabrication. Therefore, part of our contribution is based on testing the following research aim in a realistic heavily constrained context. The specific expectations of the industrial partner are

*Corresponding author: Abdelghani.Benghanem@USherbrooke.ca

threefold and define the research aims: (1) to identify a faster yet rigorous way to achieve sensory profiling of SSVs sound based on consumers' perceptions, interpretations and their expectations, (2) to investigate the effect of loudness equalization of the sound samples on sensory profiles, and (3) to determine if using virtual participants can improve the analysis of the information contained in the original perceptive evaluations conducted on a limited sample of participants. Such approach would undoubtedly help to facilitate the integration of sound quality studies in the general acoustical engineering workflow at a lower cost, but also within a more agile and flexible paradigm to ensure quick and real impact on the work of acoustical engineers.

Subsections hereafter present a relevant state of the art, as well as some essential knowledge on sound quality, required to undertake the scientific motivations and the methodology.

1.1 Sound quality and sound signature

Blauert and Jekosch [3] define sound quality as follows: Product sound quality is a descriptor of the adequacy of the sound attached to a product. It results from judgments upon the totality of auditory characteristics of the said sound, the judgments being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation. In many experiments, researchers have investigated the acoustic sound quality of the interior noise of vehicles using perceptual evaluations and physical measurements [1, 4–6]. Several perceptual evaluations of vehicle sound studies consider a single attribute, the most popular is the annoyance. This is so because the experimenters already have an idea of which sound property is related to the sound quality they want. However, it is known that sound quality is multidimensional and the overall judgment depends on a limited set of perceptual attributes [7–9]. In addition, it is clear that each vehicle category (sport, recreational, utility, etc.) has a specific sound signature.

The sound signature can be defined as the sensory profile of a sound resulting from listening test responses giving judgments about the various perceptual attributes. This can be achieved in numerous ways. The semantic differential (SD) technique [10] still seems to be one of the most frequently used methods to investigate sounds from a perceptual point of view in different situations. Typically, for SD the concept (here sound perception/sensation) is scaled on a set of successive pairs of adjectives, such as pleasant–unpleasant, smooth–rough, loud–soft, using a 7-point rating scale [7, 11]. Often, Von Bismarck's semantic differential scales are chosen as a reference [12]. However, instead of using pairs of bipolar adjectives (e.g. dull–sharp) as in the traditional differential semantic paradigm, some studies [13] rather use an attribute and its negation (e.g. sharp–not sharp) as proposed by Kendall and Carterette [14].

Another method, called multidimensional scaling has been applied to several sound quality studies. In this

technique, participants are asked to evaluate a series of sound attributes one after the other. This method has several advantages over the semantic differential as well: (1) there is no need to construct bipolar scales (avoiding difficulties in choosing contrasting adjectives that really belong to the same dimension) and (2) different adjectives can be used for describing the same dimension (e.g., loud and soft or near and far), thus providing data for homogeneity and reliability analyses [7].

Either way, one of the first steps is to define the attributes. Lists of perceptual attributes and terms are well defined for sound and audio evaluations but most of the time, specific sound products are studied, which therefore requires appropriate attributes [15, 16]. Unfortunately, there is a lack of standardized lists of attributes to accurately describe the sound characteristics perceived by SSV regular consumers [17, 18]. Yet, two methods might be useful to overcome this limitation: (1) consensus vocabulary development procedures or (2) individual vocabulary development procedures.

1.2 Rapid sensory profiling

Sensory profiling, initially used in the food and wine industry, focuses on the five senses and their impact on the consumer experience [19–22]. Sensory analysis methods to describe and to quantify the perceptual characteristics of audio stimuli have been actively researched in recent years. A classical sensory profile includes the following key steps: (a) firstly, the selection, training and supervision of a short panel of assessors, (b) the generation of attributes that describe the similarities and differences between products, (c) the determination and consensus on the evaluation procedure for each of the selected attributes, (d) the training for the evaluation and scaling of the selected attributes for stimuli, and (e) finally the quantitative evaluation of stimuli, in a randomized presentation. This conventional method, well defined in the literature, is considered the most reliable sensory profiling method. From a practical point of view, it is easy to use and communicate. However, the specificity of the automotive industry makes the use of the conventional profiling difficult. There are several issues that render it less effective: time consuming, expensive, need of dedicated resources, and the fact that this method is not adapted to reveal inter-individual differences. For example, to apply such an approach, it takes about 6 months with a 3-h session per week with experts, which generates astronomical costs. Thus, companies cannot afford to repeat this experience for every product or element, especially for an annual design cycle and agile revision of acoustical engineering on a regular basis. Quick methods for a quick turnaround are mandatory to ensure a full penetration of sound quality methods within the industrial sector. This is a primary motivation for this work, being able to fulfill the industrial sector pragmatical realities [21, 23, 24]. For instance, Bergeron et al. [25] applied the classical sensory profile technique to obtain a description of internal rolling noise of automobiles, using quantitative perceptual criteria. Despite the overall success of such a method for

studying vehicle sound quality, this approach has a major limitation, the long implementation time: the training of the participants and the consensus vocabulary phase require a total of 36 h.

Therefore, new sensory characterization methodologies called rapid sensory profiling methods were developed, for the food and the flavour industry, to avoid training sessions, and as a substitute to the classical method. These methodologies are less time consuming, more flexible and can be used with semi-qualified assessors and even consumers, providing sensory mapping that is very close to a classical descriptive analysis with highly qualified panels [19, 26].

Lorho [8], used a descriptive analysis method [24], to compare the perceptual characteristics of spatial improvement systems for sound reproduction on headphones. The experimenter used the individual vocabulary development approach. This approach proved to be faster compared to traditional consensus vocabulary methods, and the attribute test produced reasonable results in terms of perceptual description and algorithm discrimination. Lokki et al. [9] used a similar approach to obtain perceptual attributes (and perceptual profiles) for the evaluation and acoustic comparison of concert halls. That study demonstrated that such a method works well to assess subjective differences with respect to concert halls and seating positions in the same hall.

Kaplanis et al. [27] used a rapid sensory analysis, the flash profile [23], to examine the perceptual properties of automotive audio systems. The flash profile method requires only three sessions: (a) a first one for the generation of attributes, (b) one for the setting up of lists of attributes, and (c) a last one for the quantitative evaluation of stimuli. Flash profiling allows the perceptual experience to be evaluated using the attributes developed during the individual elicitation phase in a time-efficient manner. However, it has been shown that the free elicitation method used in the individual vocabulary process is more difficult for untrained participants than for assessors familiar with descriptive analysis methods. Lorho [8] proposed to use the grid-directory technique in the attribute elicitation process for assessors who are not familiar with descriptive analysis. For example, for the generation of verbal descriptors for quality assessment of perceived spatial perception of an audio system, Berg and Rumsey opted for the grid-directory technique [28]. Thus, this research highlighted the need for clear definitions and avoiding ambiguous attributes.

Le Bagousse et al. [29] conducted an exclusively lexical study (without listening to sound) to reduce the number of terms by classifying them into categories. Then, they integrated these categories to assess the perceived sound quality for several audio coding algorithms for digital audio data compression applications. Additional publications related to sensory analysis applied to acoustics can be found in the literature: Wankling et al. [26] research interests include the development of subjective descriptors to assess the perceived quality of low-frequency audio reproduction in small rooms. Mattila's research work [30] is related to the descriptive analysis of sound quality of transmitted speech in the context of mobile communications, using the semantic differentiation method [10].

The methodology employed for the experiment reported herein is inspired from the flash profiling approach. In this paper, this promising method is adapted to quantify and rapidly assess both the sound signature and sound quality of SSVs. This method is supposed to provide accurate results, similar to the results provided by conventional descriptive analysis, with a light implementation.

1.3 Participatory sound design

The second objective of this research is tightly intertwined with a sound design perspective. As product sound quality is defined as the adequacy of the sound attached to a product, it is dependent on user expectations and context. Sound design deals with the engineering of sound and a central concept is the communication that takes place when using the product [31]. Sound design aims to create or modify the timbre of product sounds to meet specific intentions [32].

Indeed, many sound quality studies focus on existing sounds or virtual sounds in the listening tests [33]. One question related to participatory research and participatory engineering is how the end-user or consumer can be included in the sound design process with a rigorous and scientific method. Consequently, we investigated the idea of integrating sound design profiling to classical sound quality listening tests. On this matter, it is clear from the literature that part of sound design changes is related to sound quality outcomes [34, 35]. Nevertheless, the participants (e.g., users, consumers) already have an idea of the sound profile they want. This paper reports an attempt for potential consumers to design their own ideal sound profile. For the SSV manufacturer, this is an important aspect as it will help in the further design stage for sound quality and product amelioration. More precisely, it will help in identifying the direction in which acoustical engineers should work.

1.4 Effect of global loudness equalization on sensory profiles

There is a constant debate between academics and industry, shall listening tests for vehicle sound quality be based on global loudness-equalized sound samples or not? Albeit this might seem like an obvious question, for a vehicle manufacturer interested in enhanced sound quality, it is not. Indeed, the current industrial and legislative context imposes reduced sound pressure levels (SPL) to vehicle manufacturers. Therefore, reducing the SPL is always a desirable avenue in this context. Indeed, levels are still very important factors in daily acoustical engineering in order to respect the noise legislation. The balance between SPL reduction and timber adjustment for sound quality is therefore a recurring source of debate among the practicing acoustical engineers. Accordingly, on the one hand, conducting listening tests with sounds that have not been loudness equalized and leading to a conclusion that a change in loudness/SPL is desirable would not be helpful. On the other hand, conducting listening tests with loudness-equalized sounds somehow denatures the tested sounds or products. Based on that dual

position, there is a need to investigate the effect of loudness equalization on sound quality experiments. Few hints along these lines exist in the current literature [36–38].

Parizet et al. [36] evaluated perceptual prominent factors on sound perception for noises recorded in a high-speed train, with sounds at their real levels, and then with sounds equalized in loudness. The results of this study were that: (1) the first influencing factor of interior noise perception is loudness of the signal. The influence of loudness is quite the same for every listener, and loudness is mainly due to the speed of the train and (2) when the influence of loudness is eliminated, the perception is different among listeners. Most of them (70%) prefer the noise to be in the low frequency range while some other listeners choose opposite preference. However, it should be pointed out that in this second case, the experimenters failed to develop a predictive model of preference based on existing psychoacoustic descriptors using these data. Susini et al. [37] evaluated the influence of loudness on sound recognition based on an explicit memory experiment. The results of this study revealed that recognition scores were significantly different when sounds are presented with the same or different SPL between the study phase and the test phase; recognition was significantly better when target sounds are presented with the same typical level in the study phase and in the test phase.

Since most studies suggest that loudness affects sound quality, studies in sound quality systematically equalize sound samples in global loudness to help test participants focusing on finer details rather than the obvious overall perceived loudness [39]. For instance, Kwon et al. [40] proposed a model of psychoacoustic sportiveness for vehicle interior sound excluding the effect of loudness. The model for psychoacoustic sportiveness was determined as a function of roughness, sharpness, and tonality. However, there are only a limited number of studies of vehicle sounds in the literature that compare the two approaches (i.e. loudness-equalized or not) [38]. Besides these works, the question remains open for SSVs. All in all, considering the fact that the different methods and specific types of sounds condition used in each study affect the results, this needs more investigation.

Therefore, an experiment was set up to prove to SSVs manufacturers that the global sound equalization of sounds affects the sensory profiles and this can be demonstrated more precisely with two listening test sessions based on the same sound samples and exact same methodology (including hardware, participants, and conditions). To do so, in the first test, sound samples were presented at their real loudness levels, and in the second test, the global loudness of sound samples was equalized.

1.5 Bootstrap

The fourth objective of this research is related to the size of available data and the required time for sound quality studies. Testing sounds with actual customers or experts takes time and resources. One of the scientific questions is how the analysis of results can be enhanced using a method to simulate a pool of virtual participants based on a limited

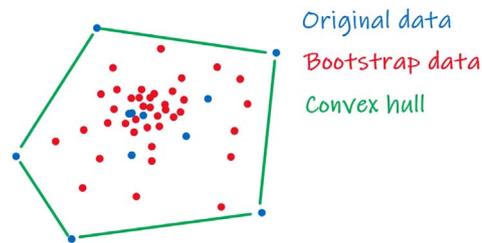


Figure 1. Conceptual illustration of bootstrapped data in two-dimension space. The boot-strapped points (red) lie within (or on) the convex hull (green) of the original data (blue).

pool of actual participants [41–43]. In this trend, we investigated the application of the bootstrapping methods [41] to sound quality studies. To the authors’ knowledge, this has not been examined for sound quality studies. The question then being: does bootstrapping facilitate the interpretation of the results and does it agree with results obtained from more established methods?

Bootstrapping is a family of methods based on random sampling with replacement of a given data set to virtually increase the number of observations. In other words, bootstrap approaches are likely to generate data similar to the observed data. For instance, in this study, assuming the case of N observations (listeners) of A attributes for the sound sample i and condition b stored in a data matrix $\mathbf{X}_{b,i} \in \mathbb{R}^{N \times A}$, bootstrapping would involve M re-sampling lines of $\mathbf{X}_{b,i}$ with replacement leading to an augmented data matrix $\mathbf{X}'_{b,i} \in \mathbb{R}^{M \times A}$. Several properties of the bootstrapping method in relation with this work and the implemented procedure are worth mentioning. First, bootstrapping will lead to mean values equivalent to the mean of the original data. Second, the confidence interval on the statistical estimates will be smaller. Third, the bootstrapping re-sampled points in a A -dimension space will lie within the A -dimension convex hull [44] of the original data set (see Fig. 1).

Indeed, the implementation of the bootstrap is of little value when the statistical inference can be made by classical analytical methods, for which the conditions of application are met. It is therefore not intended to replace conventional statistical inference methods where these are applicable, but rather to provide answers to questions for which conventional methods are inapplicable or unavailable. However, for the principal component analysis based on the variance of data, the use of the bootstrap is a valid alternative to greatly improve the readability of the data [45].

1.6 Principal component analysis

In sensory profiling, several attributes are studied simultaneously. However, it is difficult to represent simultaneously and simply such a large number of quantitative variables in a simple graph, because the data are no longer represented in a two-dimensional space but in a larger dimensional space whose interpretation is more difficult. Moreover, it is important to retain the most meaningful uncorrelated attributes. The multivariate statistical analysis tools answers these questions and reveals the hidden

structure of the data. The principle is to project the responses of the participants in the listening tests onto the principals components (PC) dimensional space.

In this regard, the chosen statistical tool for this study is the Principal Component Analysis (PCA). The main objective of the PCA is to reduce the dimensionality of a data set consisting of a large number of interdependent variables, while maintaining as much as possible the variation present in the initial data-set. Besides reducing the number of dimensions, the PCA enables identifying insignificant or unreliable data, thus de-noising the data. Exhaustive information on PCA can be found in [46–48].

1.7 Research objectives

The aim of this work was based on the need to identify and adapt rapid sensory methods to the sound quality of SSV, using the reliable resources already available, and verify the effectiveness of this method. As the flash profiling approach seems to be easy to implement in sound quality context, this method was adapted in an attempt to quickly assess sensory profiles of SSV sounds from a consumers' point of view. Since participatory sound design approach seems to be less time-consuming than the regular method, this approach was used to facilitate the sound design process. Even if it has been previously investigated in the literature, it is of interest to confirm if equalizing the overall loudness reduces the correlation between scores of perceptual attributes in sensory profiles. To avoid multiple time consuming sessions, the rapid sensory method was therefore chosen to investigate this research question. The equalized and non-equalized cases are compared with the same conditions, participants, hardware and method. Finally, it was also of interest to investigate if bootstrapping can be used as a solution to overcome the limitation regarding the participants number. Thus, these points respond to the literature review, and are the contributions of this paper.

2 Method

2.1 Sound samples

Seven recreational side-by-side vehicles were considered in this study. The interior sounds of the vehicles were recorded using a binaural GRAS KEMAR Head and Torso Simulator (model 45BB) equipped with large ears and 40AD 1/2" GRAS microphones, with a sampling frequency of 48 kHz and a 24-bit resolution.

To establish the sound signature of each vehicle, we considered the following three driving conditions: (1) idle (1200 rpm), (2) constant speed (30 km/h), and (3) wide open throttle (WOT) acceleration (0–60 km/h). These conditions were chosen for different reasons. The idle condition reflects the customer's first impression of the sound signature of the parked vehicle. The constant speed condition was chosen as 95% of the operating time is at velocities below 30 km/h for these vehicles. The acceleration condition reflects the sensation associated with vehicle sportiveness.

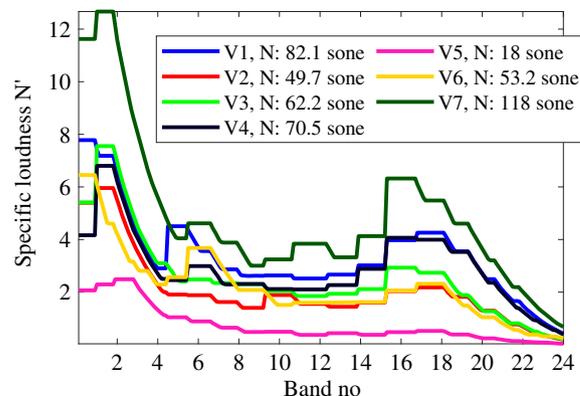


Figure 2. Specific loudness N' as a function of frequencies on the Bark scale and global loudness N (in the legend) of the seven vehicles for the left binaural channel for the constant speed condition.

Next, all sound samples of each driving condition were global-loudness equalized approximately to 29.3 sone for the idle condition, 49.9 sone for the constant speed, and 48.9 sone for the acceleration.

To illustrate the loudness differences, an example of the specific and global loudness for the seven sound samples of the constant speed condition is shown in Figure 2. In this case, the sound samples were equalized in global loudness in comparison to the global loudness of the V2 sample (red line). The global loudness values were calculated according to the ISO532B model. The equalization was applied iteratively by applying a linear gain on the sound to be equalized. This gain adjustment allows the global loudness (the area under these graphs) to be the same as for the red line with a threshold of 0.25 sone. This equalization is done for both channels (left and right) in a single process with a single gain (same for left and right, in order to preserve the binaural sound image).

As a result, for each driving condition, two versions of sound stimuli were produced. Using these stimuli, two similar perceptual test sessions were organized with the same panel of participants: a first session with sounds played back at actual sound pressure levels, and a second session with sounds with the same global loudness.

Sound stimuli were presented to participants on SENNHEISER headphones (chosen among models HD600, HD555, HD598, or HD579). Each headphone was frequency-equalized by filtering each sound sample with appropriate frequency response amplitude-only for the left and right channels of each headset, using 2048-order zero-phase finite impulse response filters. This creates the same output signals from the headphones as those measured with the binaural mannequin. Then, short fade-ins and fade-outs were applied in the beginning and at the end of each sound sample so that they could be repeated without audible artifacts. The total duration of each sample was 5 s. A validation of sound reconstruction was performed: the binaural microphones were installed on the KEMAR mannequin (equipped with the same artificial ears specified above) with the binaural headsets and the spectrum of the sounds

recorded on the mannequin was compared with the spectrum of the original sounds measured in the interior of the vehicle cabin in operating condition.

2.2 Participants

A jury of 20 individuals (17 males, 3 females) between 23 and 54 years of age participated to the study. All participants were users of recreational products similar to the studied SSVs and all of them had hearing thresholds below 25 dB HL (hearing level). The participants' hearing thresholds have been measured using an online hearing test¹ before the beginning of the listening tests. This study was reviewed and approved by the Comité d'éthique pour la recherche, the Internal review Board at Université de Sherbrooke, Québec, Canada. Informed consent was obtained from all participants before they were enrolled in the study. Participants did not receive financial compensation for their participation in the tests.

2.3 Step #1: Group discussion and attribute identification

The aim of the group discussion was to define the attributes associated with the seven recreational vehicles according to sound perception. The discussion took place in a large room in the presence of all participants and members of the research team. In this phase, monophonic sounds were presented using a Genelec 8040B loudspeaker. All the group discussions were held in French, as the target language for attributes was also French. This experience was structured in two tasks.

2.3.1 Generation of attributes

The first task aimed at generating the perceptual attributes. A preliminary training and familiarization session was included because some of the participants were not familiar with sensory profiling. Sounds were played randomly and participants freely proposed terms describing the various sounds. The duration of this training was about 30 min.

Then, participants were asked to freely describe the sound samples using verbal descriptors by entering them in an online form. The sound samples were randomly presented (the original sounds and the sounds equalized in global loudness, have been mixed to form the random playlist of sounds), several times with breaks, until each participant was able to generate about 10 attributes, which took about 45 min to complete. At the end of this phase, a total of 255 terms were collected. Then, all the descriptors generated by all the participants were compiled, sorted in decreasing order of occurrence, and then shown to the panel on a large screen in front of the room.

2.3.2 Selection of attributes

The second task aimed at selecting the most relevant descriptors. Based on a group discussion, antonyms and

synonyms were eliminated among the chosen descriptors and those deemed the most important were selected among the remaining descriptors: Firstly, the panel grouped similar words together, e.g. Power with Powerful and Vibration with Vibrating. Then, all words were quickly examined to group synonyms and antonyms together, in a simple way according to the panel's common understanding. Here are a few examples of words grouped together: Regular synonym of Constant, Performing synonym of Powerful, and Weak antonym of Powerful. Otherwise, when the group discussion tended to diverge on a word, it was left as it was, so as not to influence individuals on their understanding of the words. For example, bass versus high-pitched. We have to keep in mind that participants are just SSV users and not experts in sensory analysis. For most of them, this is the first time they were participating to such a study. Therefore, they probably do not have a vocabulary as developed as the experts or the acousticians. It is important to note that the group discussion was open and inclusive and was managed by three experimenters, one of which coming from marketing with experience in managing focus groups. A total of 128 terms were kept at the end of this task that took approximately 1 h. The perceptual attributes obtained from the group discussion are presented in Table 1. This table shows the most recurrent descriptors (occurrence larger than 1) generated by the jury panel with the occurrence of each word. Note that all the tests were conducted in French. For this article, the attributes have been translated into English using the Cambridge French–English Dictionary.

Since one participant miss-understood the instructions and made analogies with food terms, some of the terms in the footnotes are not related to sound.

Two individual voting sessions using anonymous online survey were then conducted. In the first vote, each participant had to select a list of 10 attributes from the list defined previously, which they considered relevant. These 20 attributes are assumed to be commonly used and understood by the majority of subjects. Table 2 indicates the results of the first vote session.

In the second vote, each participant had to select the most important six attributes from the 10 attributes retained after the first vote. Table 3 indicates the results of the second vote session. The six attributes selected by the jury panel at the end of the two voting sessions are shown in gray. At the end, the jury agreed in a consensus on this short list of six sensory attributes for the quantitative evaluation of SSV sounds. The attributes of this selection were then defined as the perceptual dimensions to be evaluated in the listening tests.

The two voting sessions and final consensus on six attributes took 30 min. Flash profiling is normally based on the use of free choice profiling that allows participants to use their own list of attributes. In order to facilitate the later analysis of the results, this step was slightly modified to generate a common list of perceptual attributes, compared to the standard flash profile method. The two individual voting sessions with a brief group discussion, will make it possible to have well-defined and common attributes for the entire jury panel.

¹ <https://hearingtest.online/>

Table 1. The most recurrent French attributes (31 words). The six attributes selected in the final list are formatted in bold.

Attributes	English translation	Occurrence
Puissant	Powerful	20
<i>Strident</i>	Strident	14
Doux	Soft	13
<i>Sourd</i>	Deaf	9
Métallique	Metallic	9
Vibrant	Vibrating	7
Agressif	Aggressive	6
<i>Grave</i>	Bass	6
<i>Sillement</i>	Hissing	5
<i>Étouffé</i>	Muffled	5
<i>Sec</i>	Dry, curt, sharp	5
Bruyant	Noisy	5
<i>Aigu</i>	High-pitched	3
<i>Constant</i>	Constant	3
<i>Fade</i>	Tasteless	3
<i>Frottement</i>	Friction	3
<i>Lourd</i>	Heavy	3
<i>Sifflement</i>	Whistling	3
<i>Mécanique</i>	Mechanical	3
<i>Bourdonnant</i>	Buzzing	2
<i>Essoufflé</i>	Breathless, puffed	2
<i>Étourdissant</i>	Deafening	2
<i>Feutré</i>	Felted	2
<i>Grognement</i>	Growl	2
<i>Ralenti</i>	Idle	2
<i>Onctueux</i>	Creamy, smooth	2
<i>Rapide</i>	Rapid	2
<i>Régulier</i>	Regular	2
<i>Résonant</i>	Resonant, echoing	2
<i>Saccadé</i>	Jerky	2
<i>Stimulant</i>	Stimulating	2

Descriptors occurred once (in French): *Acide, Adoucissant, Aiguiser, Aspiré, Balayeuse, Bouillonnant, Brulé, Caoutchouteux, Caverneux, Chargé, Chatouille, Claire, Commun, Compresseur, Confortable, Courroie, Crémeux, Crissement, Croissant, Croustillant, Dépassement, Destructif, Différentiel, Élégant, Éloigner, Énergique, Ennuyeux, Envahissant, Équilibré, Excitant, Expiration, Fatiguant, Filtré, Fondant, Fréquence, Froid, Gouteux, Granulaire, Gronde, Grunty, Gutturale, Huileux, Impatient, Industriel, Inégale, Insécurisant, Inspiration, Insupportable, Intense, Irritant, Juteux, Laveuse, Lointain, Louvoisement, Mélanger, Mixte, Mou, Muffler, Neutre, Oscillant, Pâteux, Pétilant, Petit, Plaisant, Prêt, Progressif, Rafale, Rampé, Râpé, Rassurant, Rauque, Réconfortant, Redondant, Retenu, Riche, Robuste, Ronronnement, Rugueux, Rythmé, Sablonneux, Salé, Sirène, Solide, Soufflant, Sportif, Stable, Subtil, Texturé, Tonnerre, Traction, Trafic, Aspirateur, Vaillant, Vent, Ventilation, Vrombissant, Wouwouwou.*

For comparison purposes, only 3 h were needed to define the perceptual attributes whereas the classical method as described in [25] requires twelve sessions of 3 h (36 h in total).

2.4 Step #2: Listening tests

In this step, each participant performed two similar listening tests: (a) with sounds not equalized in global

Table 2. First vote session results. The six attributes selected in the final list are formatted in bold.

Attributes	English translation	Votes
Puissant	Powerful	18
Vibrant	Vibrating	16
Métallique	Metallic	15
Agressif	Aggressive	14
Doux	Soft	13
<i>Mécanique</i>	Mechanical	13
Bruyant	Noisy	13
<i>Constant</i>	Constant	12
<i>Bourdonnant</i>	Buzzing	12
<i>Strident</i>	Strident	11
<i>Sourd</i>	Deaf	11
<i>Étouffé</i>	Muffled	11
<i>Grave</i>	Bass	9
<i>Sifflement</i>	Whistling	8
<i>Sillement</i>	Hissing	7
<i>Lourd</i>	Heavy	5
<i>Fade</i>	Tasteless	4
<i>Sec</i>	Dry, curt, sharp	4
<i>Frottement</i>	Friction	3
<i>Aigu</i>	High-pitched	1

Table 3. Second vote session results. The six attributes selected in the final list are formatted in bold.

Attributes	English translation	Votes
Puissant	Powerful	16
Agressif	Aggressive	14
Métallique	Metallic	12
Bruyant	Noisy	12
Doux	Soft	11
Vibrant	Vibrating	11
<i>Bourdonnant</i>	Buzzing	8
<i>Constant</i>	Constan	8
<i>Sourd</i>	Deaf	8
<i>Étouffé</i>	Muffled	7
<i>Mécanique</i>	Mechanical	5

loudness (Test NOEQ) and (b) with sounds equalized in global loudness (Test EQ). Participants had a 1-h break between the two tests. Each test was structured in three experiments with an average duration of 45 min each.

2.4.1 Experiment 1 – Sensory profile

In this first experiment, the objective was to determine the sensory profile for the sounds of the seven vehicles. Thus, listening tests were conducted to evaluate perceptual attributes using continuous intensity scales. The listening tests were set up in a quiet meeting room. During the test, the participant was seated in front of a laptop screen with a graphical user interface (GUI) used to play the sounds and to give ratings, by adjusting a slider from 0 to 100, for each perceptual attribute. Each participant was asked to rate the seven sound samples for the three driving conditions and for each of the six perceptual dimensions in Table 3. An example of the user interface for this experiment is

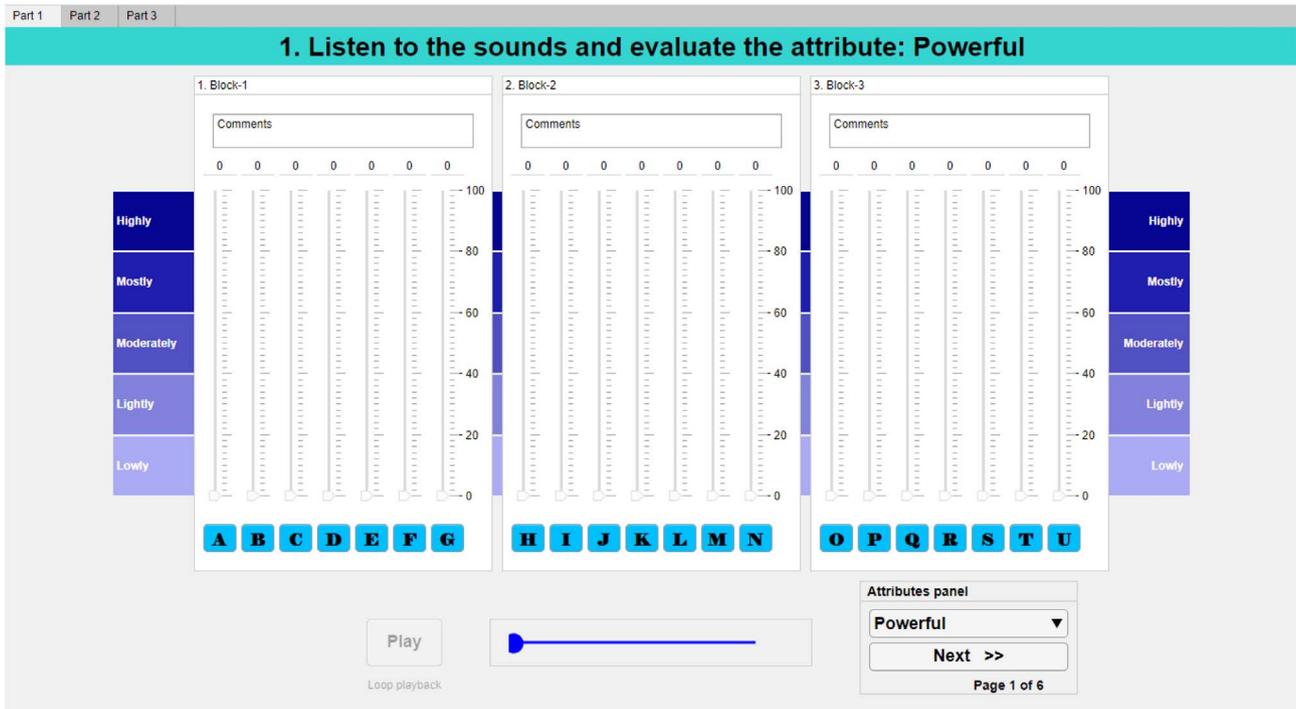


Figure 3. GUI for sound signature assessment (translated from French), presented in full-screen mode to the participants for clarity.

provided in Figure 3. Each block corresponds to a driving condition. The bold letters (A–G, H–N, and O–U) are the sound samples of the seven vehicles presented in random order in each block and for each attribute. The GUI allows the possibility to play the sounds back as many times required.

2.4.2 Experiment 2 – Evaluation of global preference

The objective of the second experiment was to rate the vehicles according to a global preference criterion. This experiment consisted in evaluating the desire-to-purchase (*Envie d’achat* in French) of the products by a sound rating test. In this experiment, the sounds were the same as those assessed in experiment 1. The participants were asked to rate the sounds of the seven tested vehicles, according to the desire-to-purchase, for each of the three conditions. The GUI used for this experiment is similar to that shown in Figure 3 except for the evaluated attribute (desire-to-purchase).

2.4.3 Experiment 3 – Participatory design of a target sound signature

The aim of the third experiment was to determine what users want as the preferred or target sound signature of a chosen reference vehicle.

In this experiment, each participant was instructed to focus on the attributes previously assessed in experiment 1 and was asked to design the “best” sensory profile for the reference vehicle (V1), according to their preferences. The initial ratings for each attribute in this test were those

assigned by the individual subject for the sound (V1) in experiment 1. Each participant gave a desired score (from –20 to 120 points) for each perceptual attribute and for each of the three driving conditions (acceleration, idle and constant speed). The scale was extended below 0 and beyond 100 to ensure that if an attribute already received 100 (or 0) in experiment 1, it was possible to go beyond (or below).

2.5 Data analysis

Before creating the sound profiles from the listening tests, a statistical analysis was performed to demonstrate the statistical significance of the responses [49].

Firstly, we performed the Shapiro–Wilk test [50] to determine if the null hypothesis of composite normality is a reasonable assumption regarding the population distribution of each sample. The returned value of $H = 0$ indicates that Shapiro–Wilk test fails to reject the null hypothesis at the 5% significance level. Thus, as the collected data did not follow a normal distribution, Friedman’s test [51] was used rather than classical ANOVA.

In this study, we considered three criteria (or three dimensions) on which Friedman’s test was performed. In each case, Friedman’s ANOVA table indicates the probability of falsely of the null hypothesis of the group on observations. The significance level was set to $\alpha = 0.05$. The three statistical criteria are as follows:

1. The group of seven vehicles: the nullity hypothesis would be that the sounds of the vehicles presented have no effect on the participants’ responses.

Table 4. Friedman’s ANOVA test results.

Criteria	p -value NOEQ	p -value EQ
Vehicles	1.78×10^{-60}	3.50×10^{-7}
Conditions	9.08×10^{-31}	3.46×10^{-6}
Attributes	1.93×10^{-7}	1.40×10^{-3}

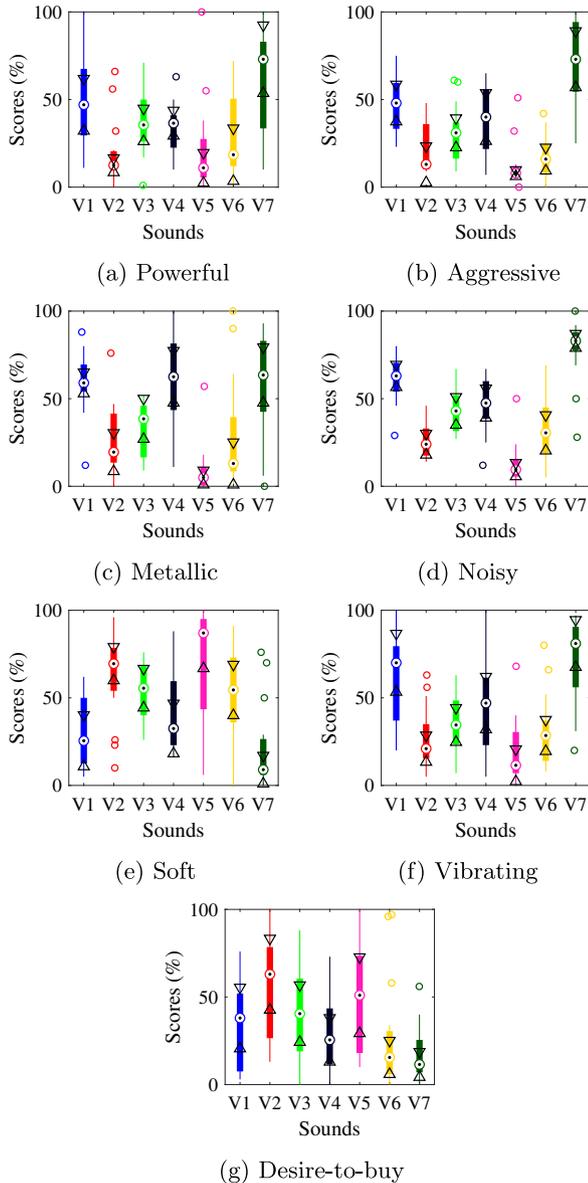


Figure 4. Box-plot of attributes rating (a–f) and desire-to-buy rating (g) for constant speed condition. Test NOEQ. The x -axis represents the seven sound samples evaluated and the y -axis represents the ratings in % of the participants for each attribute. The median is represented by a black dot surrounded by a colored circle. The 25th and 75th percentiles are presented by a thick solid line. The thin lines added at the ends extend to the extreme values (maximum and minimum) and the outliers are marked as circles. The confidence interval of the median values is represented by triangles. (a) Powerful, (b) Aggressive, (c) Metallic, (d) Noisy, (e) soft, (f) Vibrating and (g) Desire-to-buy.

2. The group of the three driving conditions: the nullity hypothesis would be that the driving conditions presented have no effect on the participants’ responses.
3. The group of the six attributes evaluated: the nullity hypothesis would be that the attributes presented have no effect on the participants’ responses.

3 Results and interpretation

3.1 ANOVA and box-plots

The p -values of the ANOVA are given in Table 4. For each criteria, the p -value is smaller than the significance level (α). This means that vehicle sounds, driving conditions, and attributes have a statistically significant effect on the response of participants.

To facilitate the reading, only the results for constant speed condition are reported in this paper, as this condition represents 95% of the operating time of these SSV.

As a representative example, Figure 4 illustrates the responses of the participants in the listening tests, as box-plots form, for the constant speed condition and sounds not equalized in global loudness. In each box-plot diagram, the scores in % are represented for each sound sample.

The box-plots for all participants show some outliers and some wide interquartile intervals. This is probably due to a lack of consensus in consumer responses. Also, from Figure 4, comparisons can be made between the sound profiles of the seven vehicles. For the constant speed condition, the sound of V5 is evaluated the least aggressive, the least noisy, the softest, the least metallic, the least powerful and the least vibrating. Moreover, the sound of V7 is evaluated the most aggressive, the noisiest, the less soft, the most metallic, the most powerful and the most vibrating.

Figure 4g presents the box-plot diagram of the evaluation of the vehicle sounds according to desire-to-buy (the second experiment of the listening test) for the constant speed condition. For this condition, Figure 4g shows that V2 is the preferred vehicle sound in terms of desire-to-buy (63%).

3.2 Sensory profiles

In this section, the current and desired sound profiles of the tested vehicles are presented and compared. The sound signature or sound profile of each vehicle is illustrated according to the six attributes as a radar plot. The median values of the participants’ responses for each attribute were used to design these sound profiles.

The results are displayed for original sounds without global loudness equalization (Test NOEQ) and for the sounds with global loudness equalization (Test EQ). Figure 5 presents the current sound profiles of the seven vehicles tested at constant speed.

Figure 5 shows that the sensory profiles are different for the different vehicles. This suggests that the assessors were able to discriminate between the stimuli. Also, Figure 5 shows that the sensory profiles for original sounds (solid lines) are different to those for the sounds with global

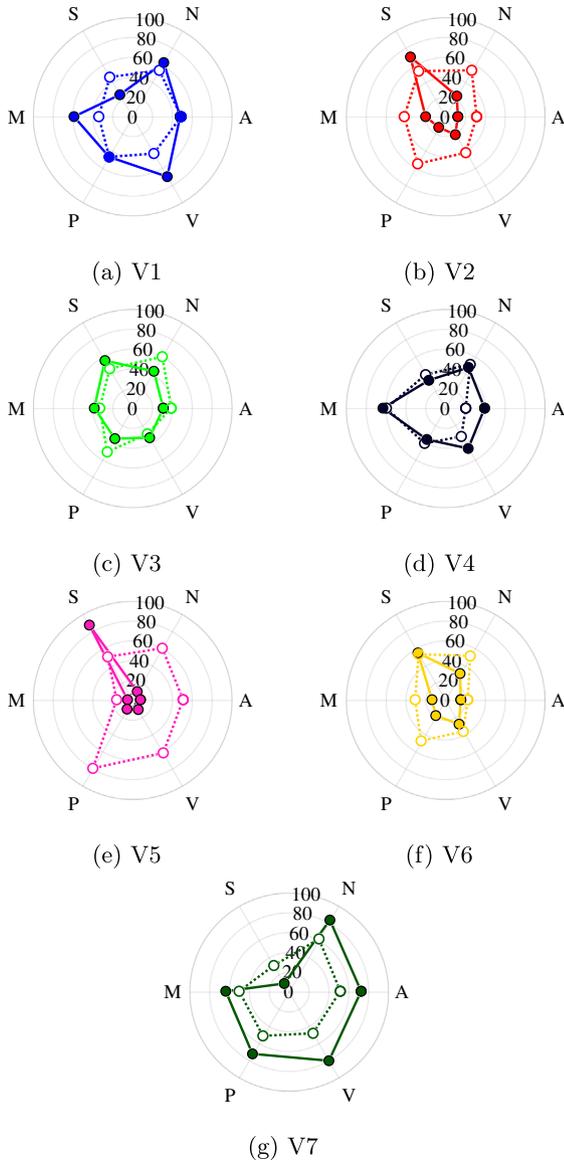


Figure 5. Sensory profiles for constant-speed condition. Solid lines: Test NOEQ. Dotted Lines: Test EQ. A: Aggressive, N: Noisy, S: Soft, M: Metallic, P: Powerful, V: Vibrating. (a) V1, (b) V2, (c) V3, (d) V4, (e) V5, (f) V6 and (g) V7.

loudness equalization (dotted lines) except for V3 and V4 that have close profiles. The NOEQ profiles show large variations of the Noisy and Soft attributes among vehicles (consider V5, V6, V7). This is expected since Noisy and Soft are likely to have a large correlation with global loudness. In contrast, the EQ profiles show more balanced scores of all six attributes. This suggests that the global loudness equalization of the sound stimuli has an effect on subjective assessments of perceptual attributes and thus on sensory profiles.

3.3 Participatory sound design

Figure 6 presents the sound profile for the reference vehicle (V1) and the desired sound profile by the panel of

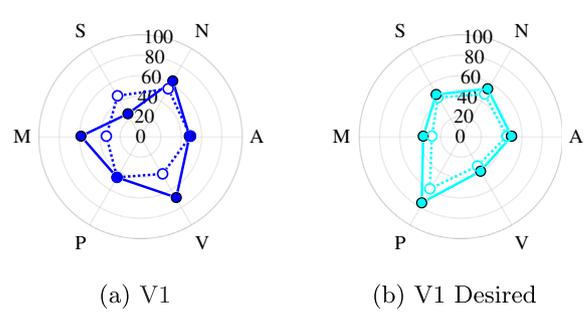


Figure 6. Sensory profiles “desired” for V1 for constant-speed condition. Solid lines: Test NOEQ. Dotted lines: Test EQ. A: Aggressive, N: Noisy, S: Soft, M: Metallic, P: Powerful, V: Vibrating. (a) V1 and (b) V1 desired.

assessors. Firstly, Figure 6 shows that the sensory profiles desired by assessors for the reference vehicle (V1) were very similar for EQ and NOEQ results. This is a very interesting result that support the idea that a desired sound style or signature might not so much be defined by the loudness. For instance, as an illustrative example, a well-branded car sound would sound as desirable and recognizable while listening to a YouTube video or seeing one live in the far distance, i.e., the recognizable signature should not be depend on level or perceived loudness.

Secondly, results from Figure 6b give valuable clues on how to improve the sound signature of V1, by comparison with Figure 6a: the participants want a softer, less metallic, more powerful and less vibrating sound profile of V1.

3.4 Effect of global loudness equalization on sensory profiles

This section studies the effect of global loudness equalization on sensory profiling using two listening test sessions with and without global loudness equalization. To do so, correlation analyses were performed to evaluate the relationship between the subjective assessment of perceptual attributes and the global loudness of the sound samples.

The results of these correlation analyses are presented in Figures 7–9 for the two listening tests. Figure 7 illustrates the correlations between the six perceptual attributes scores and the global loudness of each of the seven sounds. Figure 8 presents scatter plots of the attributes scores. Figure 9 presents the correlations between the desire-to-buy scores and the global loudness of each of the seven sounds.

Table 5 presents the matrix of correlation coefficients (R) and the matrix of p -values (P) for testing the hypothesis that there is no relationship between the observed phenomena (null hypothesis). 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.

Figure 7a shows that for the listening tests with sound stimuli not equalized in global loudness, all perceptual attributes scores were correlated with global loudness: the attributes Aggressive, Noisy, Metallic, Powerful and Vibrating increase with global loudness (positive dependence) while the attribute Soft decreases with global loudness (negative

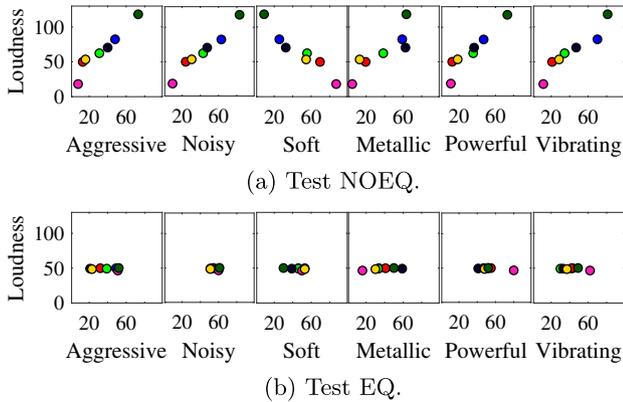


Figure 7. Scatter plots of the median scores of perceptual attributes (in %) and global loudness of sounds (in sone) for constant-speed condition. Legend: ●V1, ●V2, ●V3, ●V4, ●V5, ●V6, ●V7. (a) Test NOEQ and (b) Test EQ.

dependence). These observations are confirmed by large correlation coefficients between global loudness and attributes scores: 0.96 for Aggressive, 0.98 for Noisy, -0.96 for Soft, 0.86 for Metallic, 0.95 for Powerful, and 0.95 for Vibrating. These values were computed using the Pearson product-moment correlation coefficient. This suggests that global loudness dominates the sensory profiles, all perceptual attributes being proportional, or inversely proportional, to global loudness of the sounds. In contrast, [Figure 7b](#) shows that for the tests with sound stimuli equalized in global loudness, listeners were able to perceive and

quantify differences in perceptual attributes scores for different vehicles with the exception of the attributes Noisy and, to some extent, Soft, which show small score dispersion. This is expected since these perceptual attributes are directly correlated to global loudness. It can therefore be assumed that global loudness equalization forces the listeners to do a finer analysis of the perceptual attributes. The score dispersion for the Aggressive, Metallic, Powerful, Vibrating attributes is somewhat smaller for the EQ test as compared to the NOEQ test, which shows that vehicle sounds are more difficult to discriminate for EQ sounds with respect to those attributes. Also, the ranking of vehicles with respect to those Aggressive, Metallic, Powerful, Vibrating attributes is quite different for the EQ and NOEQ tests, which shows that global loudness masks the evaluation of less dominant attributes.

The scatter plots in [Figure 8a](#), show that the perceptual attributes are strongly correlated with each other in NOEQ tests, which is expectable since the perceptual attributes scores are also strongly correlated with global loudness. Each attribute has a positive correlation with all other attributes except for the attribute Soft which has a negative correlation with all attributes. This observation was confirmed by the correlation coefficients computed between the perceptual attributes using the Pearson product-moment correlation coefficient. Additionally, the results in [Figure 8b](#) show that there is less or no correlation between perceptual attributes when the sound samples have been equalized in global loudness, with some exceptions. For instance, significant correlations can be found when

Table 5. Matrix of correlation coefficients (R) and p -values (P).

		Aggressive	Noisy	Soft	Metallic	Powerful	Vibrating
(a) Test NOEQ							
Aggressive	R	1					
	P	1					
Noisy	R	0.987	1				
	P	0.000	1				
Soft	R	-0.949	-0.971	1			
	P	0.001	0.000	1			
Metallic	R	0.905	0.897	-0.915	1		
	P	0.005	0.006	0.004	1		
Powerful	R	0.993	0.979	-0.926	0.867	1	
	P	0.000	0.000	0.003	0.011	1	
Vibrating	R	0.971	0.983	-0.970	0.894	0.956	1
	P	0.000	0.000	0.000	0.007	0.001	1
(b) Test EQ							
Aggressive	R	1					
	P	1					
Noisy	R	0.800	1				
	P	0.031	1				
Soft	R	-0.304	-0.360	1			
	P	0.508	0.427	1			
Metallic	R	-0.381	-0.302	-0.662	1		
	P	0.399	0.510	0.105	1		
Powerful	R	0.542	0.573	0.302	-0.736	1	
	P	0.209	0.179	0.510	0.059	1	
Vibrating	R	0.687	0.477	-0.005	-0.493	0.842	1
	P	0.088	0.279	0.992	0.261	0.017	1

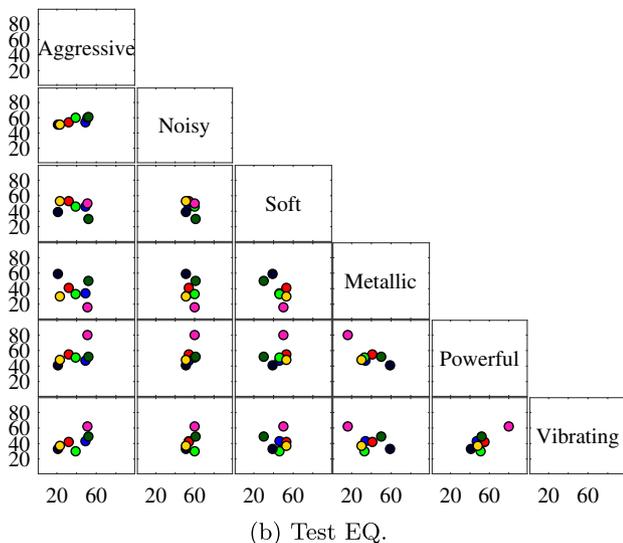
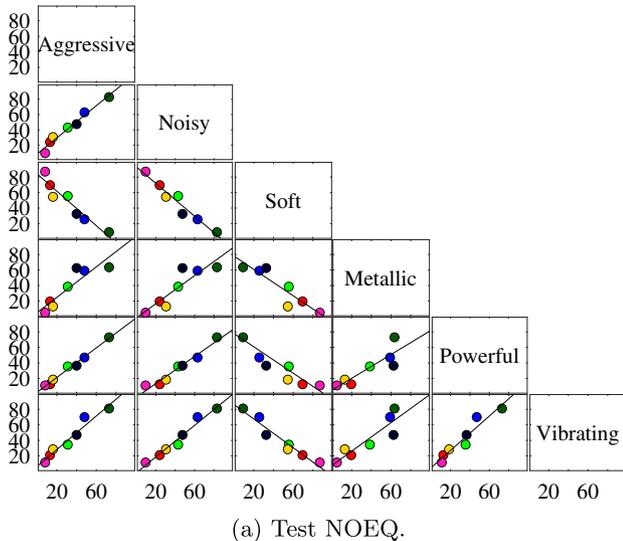


Figure 8. Scatter plots of perceptual attributes for constant-speed condition. The scatter plots correspond to the median scores (% for each attribute) for each of the seven sounds. The variable names displayed along the diagonal of the matrix are also the column names for labeling the x and y axes. For instance, the top left graph is a comparison between Aggressive and Noisy, and so on. Legend: ●V1, ●V2, ●V3, ●V4, ●V5, ●V6, ●V7. (a) Test NOEQ and (b) Test EQ.

comparing Aggressive and Noisy (0.8), or Powerful and Vibrating (0.84): this suggests that in EQ tests, participants did not make a large difference between these pairs of perceptual attributes, in these specific cases. Indeed, the strong correlation between Powerful and Vibrating in this study is mainly due to one stimulus (V5) which is considered the most Powerful and most Vibrating distantly from the other 6 stimuli. This is probably due to the particular timbre of this sound.

Finally, Figure 9a show that the desire-to-buy is strongly correlated to the global loudness level: the desire-to-buy scores decreases with the global loudness (negative dependence).

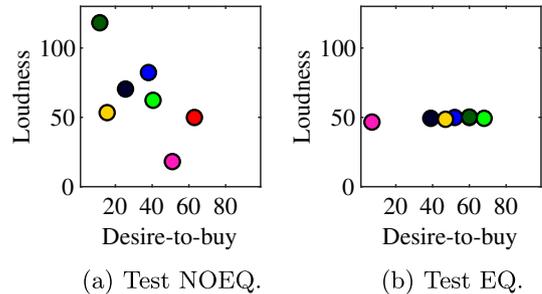


Figure 9. Scatter plots of the desire-to-buy attribute for constant speed. Legend: ●V1, ●V2, ●V3, ●V4, ●V5, ●V6, ●V7. (a) Test NOEQ and (b) Test EQ.

These results confirm the strong dependence of the subjective evaluation of the sound samples on their global loudness level and suggest that global loudness equalization reduces the correlation between perceptual attributes scores of sensory profiles for the tested vehicles. In addition, based on the dispersion of the perceptual attributes scores, it can be assumed that global loudness equalization forces listeners to more finely differentiate the perceptual attributes. It should then allow for a more detailed sensory profile differentiation. These outcomes are solidified by PCA results in Section 3.6.

3.5 Bootstrapping for creating virtual participants

The above results have been obtained on a limited number of participants (16 for NOEQ and 19 for EQ). From the 20 subjects who participated in the study, one had to leave for personal reasons and three others did not answer the full test, so their answers were not considered.

The nature of sensory profiling methodology makes it complicated, time-consuming, and costly to collect data on larger groups.

In this paper, we rely on a simple bootstrapping procedure. First, for each condition b and each sound sample i , one starts with the original sensory profile matrix $\mathbf{X}_{b,i} \in \mathbb{R}^{16 \times 6}$ in the NOEQ case ($\mathbf{X}_{b,i} \in \mathbb{R}^{19 \times 6}$ in the EQ case). Second, $\mathbf{X}_{b,i}$ is line-resampled uniformly at random sampling with replacement to obtain a new data matrix $\mathbf{X}'_{b,i} \in \mathbb{R}^{16 \times 6}$ for the l th resampling of $\mathbf{X}_{b,i}$. Note that in our case the resampled data includes the $N = 16$ original participants in NOEQ case (or $N = 19$ original participants in EQ case). From this new data matrix $\mathbf{X}'_{b,i}$ we compute the mean along each column leading to a new virtual participant that gives a sensory profile $s'_i \in \mathbb{R}^{1 \times 6}$ (which is a response of the new virtual participant for sound sample i). This is repeated L times for the index l leading to a new pool of L virtual participants. In the end, for a given condition and a given sound sample, the resulting sensory profile bootstrapped matrix is given by $\mathbf{Y} \in \mathbb{R}^{L \times 6}$. This new sensory profile matrix is then used in the subsequent stages of analysis.

Figure 10 illustrates the bi-plot of assessors' evaluations for two distributed perceptual attributes, for the original data and bootstrapped data. The number of bootstrap samples chosen for this analysis is $L = 1000$ samples. After

about 50 iterations, the results converge towards the same scores distribution. Therefore, this choice ($L = 1000$) was considered sufficient for a robust estimate. However, the choice of the number of bootstrap samples depends mainly on the size of the initial samples and the type of estimate sought. Thus, an optimal choice of number of bootstrap samples is not trivial and is often carried out by iterative or adaptive approaches.

By comparing [Figures 10a](#) and [10b](#), bootstrapping provides a better transcription of the information contained in the original data. Based on discussions with the industrial partner, it also seems that the bootstrap results are more easy to grasp and communicate. So, for sound quality and sound signature studies with a limited number of participants, the bootstrap can be used to create virtual participants.

These bootstrapped data are used in the following analysis by principal components.

3.6 Principal component analysis

This section presents a principal component analysis, applied to the responses of the listening tests to transform the correlated attributes into a few uncorrelated dimensions.

[Table 6](#) presents the eigenvalues of the covariance matrix of attributes scores, the variance values explained (EV) in % and the cumulative variance values (inertia) in %.

[Table 6a](#) indicates that 87.85% of the information contained in the data-set can be explained by a single PC, in the NOEQ case. In other words, more than 87% of the attributes scores can be ordered according to the PC1 axis and less than 13% according to the other five PC axes. This observation is coherent with the initial hypothesis that global loudness dominates the sound signature in the NOEQ test. On the other hand, for the EQ test (see [Tab. 6b](#)), the first PC axis explains 45.37% of the information, PC2 explains 23.49%, and PC3 explains 12.43%. Therefore, there are two or three important principal components for this second case, which are discussed below.

The loadings define what a principal component represents in a data-set. These are the weights (coefficients) that define a latent variable to represent a mixture of variables. Consequently, each principal component is represented by a linear combination of perceptual attributes [52] and the PCs are therefore the principal axes of the original data-set.

[Table 7](#) presents the loadings of the first three PCs (PC1, PC2 and PC3) for constant speed condition in the NOEQ and EQ tests.

[Figure 11](#) shows the loadings in the plane PC1–PC2 in the NOEQ and EQ conditions. The points indicate the weights (loadings) of each perceptual attribute in the linear combination to construct each of the PC axes. [Figure 11](#) illustrates the hidden perceptual dimensions that have permitted to distinguish the sound profiles and emphasizes the meaning of each PC axis.

According to [Figure 11a](#), the attributes Aggressive, Noisy and Vibrating are grouped together, indicating that they are correlated in the PC1–PC2 plane. The PC1 axis

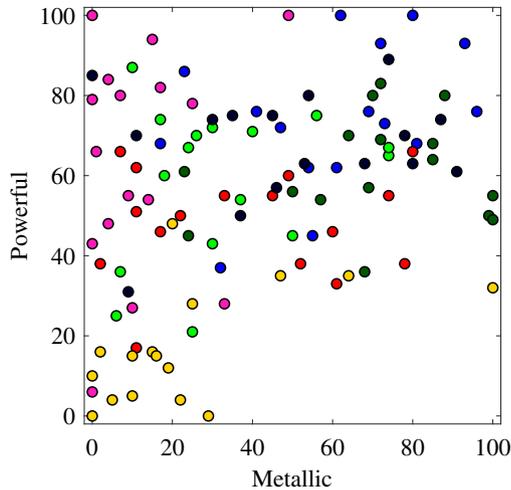
is defined by the linear combination of the attributes: Metallic, Noisy, Vibrating, Aggressive and Powerful with loadings of about -0.4 and the attribute Soft with a loading of 0.4 (see [Tab. 7a](#)). This combination is consistent with the trends observed in the analysis of correlations between perceptual attributes and global loudness of sound samples: all attributes are positively correlated with global loudness except the attribute Soft which is negatively correlated. The first principal component for the NOEQ condition is thus essentially the global loudness of sound samples. The PC2 axis in the NOEQ condition is essentially controlled by the attributes Metallic (loading -0.78) and Powerful (loading 0.58), the coefficients of the Noisy, Vibrating and Soft attributes being smaller than 0.2 .

As [Figure 11b](#) suggests, the perceptual attributes are less correlated in the EQ test. For PC1, the largest variability was observed between the attribute Metallic (positive dependence) and the group of attributes: Powerful, Vibrating, Aggressive and Noisy (with a negative dependence). It is interesting to note that PC2 of the NOEQ condition has loadings resembling PC1 of the EQ condition (with a sign inversion, and with the attributes Aggressive, Noisy and Vibrating being somewhat more important in PC1 of the EQ test). This observation tends to show that the EQ condition would essentially remove global loudness as a dominant PC and preserve the structure of the subsequent PC. For PC2 of the EQ condition, the largest variability was observed essentially between attributes Soft and Aggressive/Metallic.

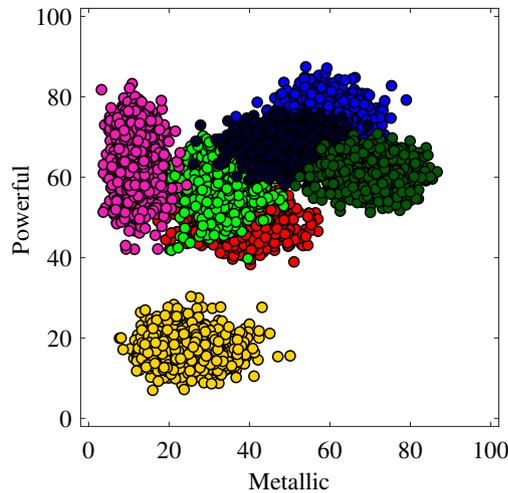
The scores represent each bootstrapped sample of the actual data projected on the PC axes. So, each score will explain some of the variation in the data-set. Therefore, the samples that are close are similar in terms of their representation on the PC axes, defined by the vectors of loadings. This suggests that it is possible to classify the sound profiles using these measured latent variables [52].

[Figures 12](#) and [13](#) present the scores in the planes PC1–PC2 and PC1–PC3, respectively for the NOEQ and EQ case. The other three PC axes are not presented since they were considered insignificant.

[Figure 12](#) shows that the distribution of sound profiles on the PC1–PC2 map is the most significant. The responses are grouped by vehicle with a high dispersion along the PC1 and PC2 axes. Indeed, sound profiles are well separated by forming distinct clusters, especially in the NOEQ case. For example, in [Figure 12a](#) the scores on PC1 shows that vehicles V7, V1 and V4 have negative values. Since PC1 in the NOEQ condition essentially predicts the opposite of global loudness, the ranking of the vehicles according to PC1 is in a decreasing order of global loudness. In contrast, in [Figure 13](#), sound profiles are grouped by vehicle but not separated along the PC3. Thus, this representation suggests that sound profiles are not significantly separated according to the third PC, and the other PCs of lower eigenvalues. These observations lead to the assumption that the significant principal components for the constant speed condition are probably the two first PCs (PC1 and PC2): these two dimensions represent about 93% of the model inertia for the NOEQ case and about 69% for the EQ case.



(a) Original data (16 observations, 7 vehicles).



(b) Bootstrapped data (1000 observations, 7 vehicles).

Figure 10. Bootstrapping in 2D: bi-plot of data for two perceptual attributes ratings. Legend: ● V1, ● V2, ● V3, ● V4, ● V5, ● V6, ● V7. (a) Original data (16 observations, 7 vehicles). (b) Bootstrapped data (1000 observations, 7 vehicles).

3.6.1 Reconstruction of sound profiles based on PCA

The previous results of the EQ test for constant speed show that the first 2 PCs predict 69% of the total variance in the initial data set, the other four PCs being much less meaningful. In this section, the original attributes ratings are reconstructed from the PC1 and PC2 scores and loadings only.

Based on the obtained PCA model, the input matrix can be retrieved by simply reversing the problem to find the data.

The results of attributes ratings reconstruction from the two first PCs are presented in Figure 14 as box-plots diagrams. These box-plots are compared to the initial attributes ratings (Fig. 4) in the same figure. The median is represented by a black dot surrounded by a colored circle. The 25th and 75th percentiles are presented by a thick solid line.

Table 6. Eigenvalues of the covariance matrix of attributes scores.

PC	Eigenvalue	EV (%)	Inertia (%)
(a) Test NOEQ			
PC1	5.2709	87.8483	87.8483
PC2	0.3086	5.1431	92.9915
PC3	0.1456	2.4272	95.4187
PC4	0.1291	2.1518	97.5705
PC5	0.0897	1.495	99.0656
PC6	0.0561	0.9344	100
(b) Test EQ			
PC1	2.7221	45.3688	45.3688
PC2	1.4097	23.4951	68.8639
PC3	0.7459	12.432	81.2962
PC4	0.4604	7.6734	88.9695
PC5	0.3723	6.2046	95.1741
PC6	0.2896	4.8259	100

Table 7. Loadings on PC1, PC2 and PC3 for constant speed condition.

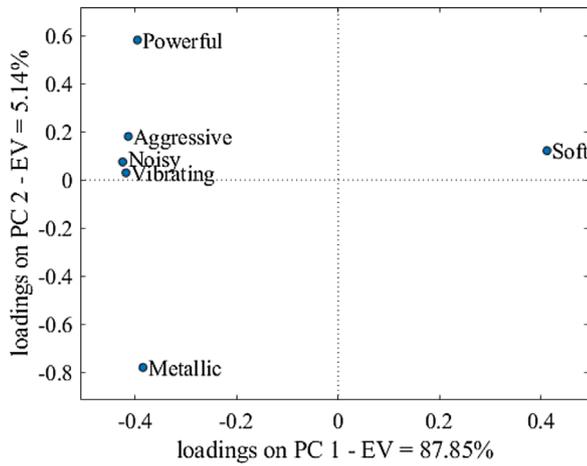
	PC1	PC2	PC3
(a) Test NOEQ			
Aggressive	-0.4137	0.1809	-0.5180
Noisy	-0.4247	0.0750	-0.0855
Soft	0.4114	0.1215	0.1246
Metallic	-0.3846	-0.7791	0.3927
Powerful	-0.3956	0.5822	0.6865
Vibrating	-0.4181	0.0303	-0.2887
(b) Test EQ			
Aggressive	-0.4211	0.4309	-0.0879
Noisy	-0.3672	-0.1883	-0.8506
Soft	-0.1148	-0.7408	0.2723
Metallic	0.4526	0.3904	-0.0372
Powerful	-0.5130	0.0479	0.3358
Vibrating	-0.4547	0.2745	0.2836

The thin lines added at the ends extend to the extreme values (maximum and minimum) and the outliers are marked as circles. The confidence interval of the median values is represented by triangles.

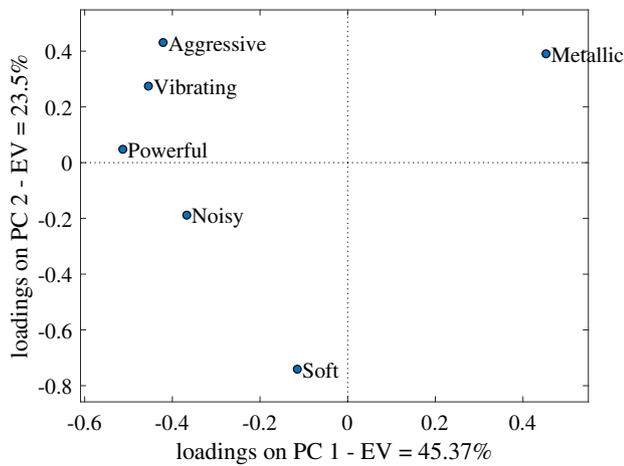
Figure 14 shows that the 25th and 75th percentiles (thick solid lines) and confidence interval of the median value (triangles) are much closer to the median. So, discarding meaningless PCs is a way to de-noise data. However, some of the reconstructed attributes seem to show a larger correlation, like Aggressive, Powerful and Vibrating. This is expected since reducing the data-set to only two independent PCs automatically creates linear dependence among some of the six attributes.

4 Discussion

This study has provided a number of experimental results about sensory profiling of SSV sounds. A rapid sensory profiling methodology, inspired from flash profiling,



(a) Test NOEQ.

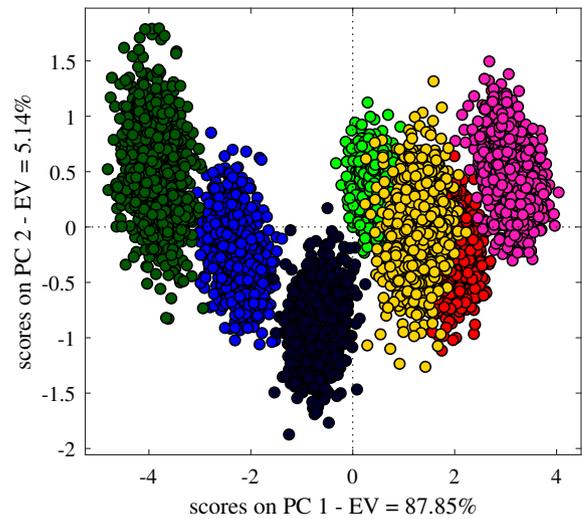


(b) Test EQ.

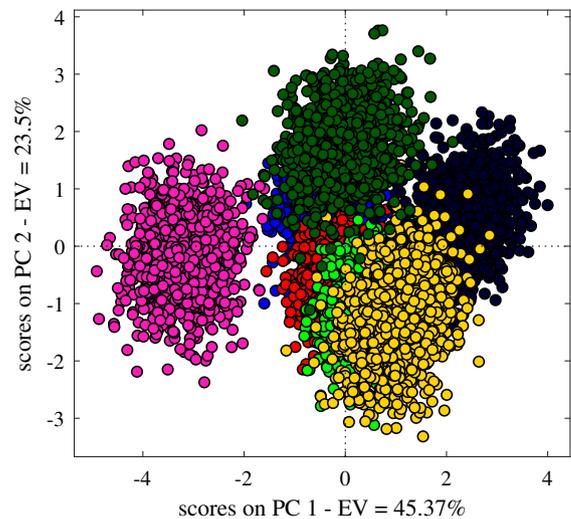
Figure 11. Loadings on PC1–PC2 for constant speed condition. (a) Test NOEQ and (b) Test EQ.

was proposed to quantify and evaluate the sound profiles, sound quality, and participatory sound design for developing a target sound of the SSV vehicles. The perceptual attributes development process employed in this study (two sessions only) showed to be very fast compared to the classical descriptive method. Thus, the proposed method is designed to provide a quick access to the sound profiles of a set of sounds. The procedure was completed in 1/8 of the time compared to the classical descriptive method. Also, the whole experiment was completed in the same day.

The first element that can potentially impact the rapid method results was the attributes generation phase. The results obtained therefore depend on the verbal skills of the participants. Without a training phase we could think that this step maybe criticized. However, we can easily see that the participants have a rich vocabulary and have covered different categories of sound terms. We could mention: (1) signal related terms (e.g. bass, high-pitched, resonant), (2) affective responses to sounds (e.g. Aggressive, Strident, Stimulating), 3) Vehicle condition related terms (e.g. rapid, constant, idle), (4) connotative associations (e.g., Powerful,



(a) Test NOEQ.



(b) Test EQ.

Figure 12. Scores on PC1–PC2 of bootstrapped data for constant speed condition. Legend: ● V1, ● V2, ● V3, ● V4, ● V5, ● V6, ● V7. (a) Test NOEQ and (b) Test EQ

Soft), (5) physical property terms (e.g. heavy, metallic, mechanical), (6) references to events and sound sources (e.g. vibrating, whistling, friction), (7) Changes in perceptions (e.g. muffled, felted), (8) direct sound descriptors (e.g. dry, jerky), and (9) onomatopoeia (wouwou). This study provided a long list of attributes that are very useful for future studies on the perception of SSV sounds.

The second critical element was the selection phase. The compilation of terms and the two anonymous voting sessions for the selection of attributes help the participants to consider terms they had previously ignored. In addition, it can be seen that the panel reached a consensus indirectly. This can be justified by the fact that apart from two terms, the top list of attributes did not really change during these steps.

Despite the fact that there are some small imperfections in the current stage (e.g. the generation of terms for

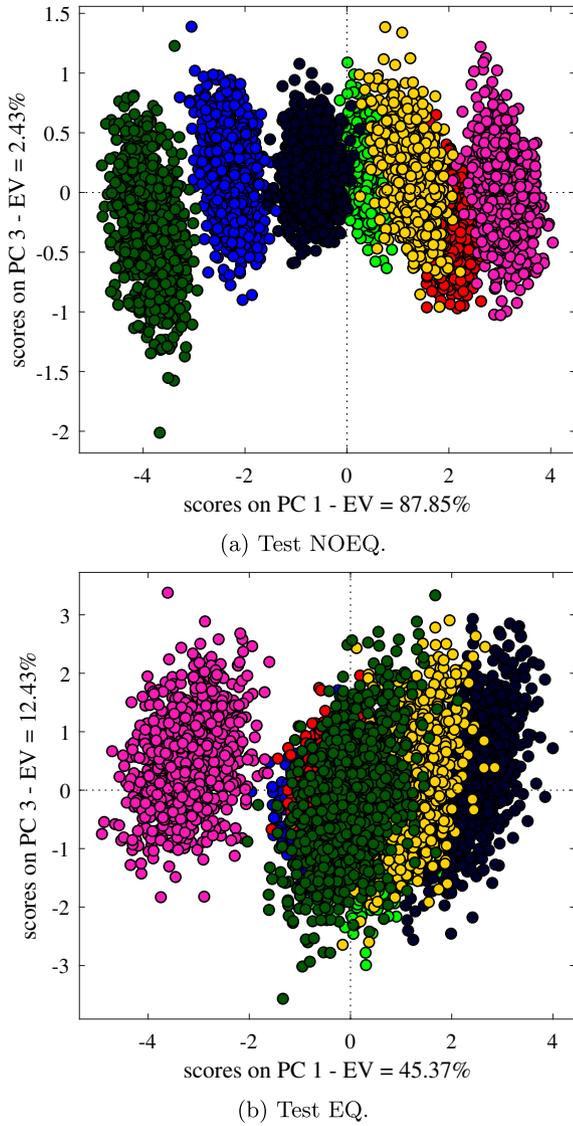


Figure 13. Scores on PC1–PC3 of bootstrapped data for constant speed condition. Legend: ● V1, ● V2, ● V3, ● V4, ● V5, ● V6, ● V7. (a) Test NOEQ and (b) Test EQ.

equalized and non-equalized sounds in global loudness at the same time, among others), this rapid procedure still shows a great potential to achieve a fast sensory profiling after the few adjustments in future studies. Furthermore, the turn-around time of the study was short enough to ensure a quick response to practical needs of the manufactures that often perform annual and iterative acoustical design.

The study showed that the participants demand a softer, less metallic, more powerful and less vibrating sound profile. Moreover, the study suggests that the “desired sound profile” might not be defined by the loudness. Having such information available in the same phase of jury testing will greatly help NVH engineers to consider the expectations of consumers and to develop this specific sound target. Overall, the participatory approach in sound design could be useful in order to rapidly improve the sound quality expected by consumers.

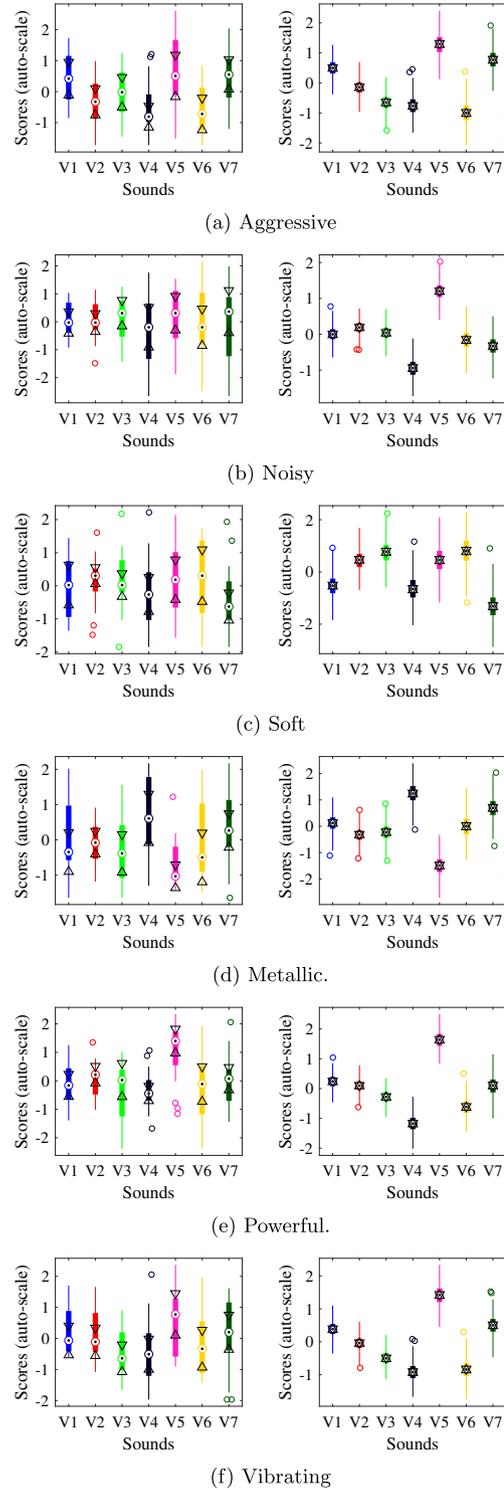


Figure 14. Box-plot of attributes ratings for constant speed condition. Left: original data. Right: data reconstructed from the first 2 PCs of bootstrapped data. (a) Aggressive, (b) Noisy, (c) Soft, (d) Metallic, (e) Powerful and (f) Vibrating.

Unsurprisingly, global loudness equalization of SSV sounds showed that the perceptual attributes were highly dependent on each other and correlate well with global loudness. This confirms the important dependency of the

subjective evaluations of sound samples on their global loudness level. Therefore it supports knowledge from the literature and also validates the overall testing reported herein.

The limited number of participants in the jury testing and the fact that the responses do not necessarily have a Gaussian distribution give the bootstrap a major advantage. The bootstrapped data provide a better transcription of the information contained in the original data. In this study, it has been applied in the simplest possible way, i.e. to simulate the distribution of assessors' responses, before performing the PCA. But, in the future, it can be mathematically adapted to compute the estimator of interest in sound quality studies based on a limited number of participants.

The PCA made it possible to see how the responses were mapped to the space defined by the PC axes. The scores of the responses of the NOEQ test were grouped by vehicle and well separated, especially on the PC1 axis (87.85% of variance explained). This axis, considered as the most perceptual dimension of these data, can be described as the decreasing order of global loudness. This suggests that participants have potentially ordered the samples based on loudness, even though they have used some other attributes. The scores of the responses for the EQ test were grouped by vehicle but less separated on the first two PCs, however we can still distinguish some different groups. Additionally, it seems that V5 has a specific characteristic sound since it stands out of the map. The result of this study indicates that two perceptual dimensions have been identified. The first dimension (PC1) can be considered as the perceptual dimension of Powerfulness of the vehicle (Metallic/Powerful). The attribute Metallic seems to be related to a negative connotation, so can be considered as the opposite of Powerfulness. The second dimension (PC) can be seen as the perceptual dimension of softness of sound (Aggressive/Soft).

Finally, the approach proposed for this study consisted in combining a rapid sensory profiling method, participatory sound design, bootstrapping, and principal component analysis. The results allowed to quickly investigate sound quality and sound signature and to identify the differences between SSV sounds. Such an approach permits to identify the few most important parameters to enhance the sound quality of a product in a short periods, which meets industrial needs.

5 Conclusion

In this study, sound quality, sound signature, and sound design of side-by-side vehicles were investigated. The study has been performed in a realistic scenario for real-life application with limited resources. Firstly, current sound profiles of seven vehicles at three driving conditions and at two global loudness conditions were obtained from a subjective assessments of six perceptual attributes. Also, a desired sound profile of a chosen reference vehicle was freely designed by participants.

The proposed rapid sensory profiling methodology, inspired from the flash profile is a fast way to understanding and anticipating consumer expectations.

Secondly, a correlation analysis between the global loudness of sound samples and the perceptual attributes was performed. The results presented in this paper showed that perceptual attributes were highly correlated with global loudness of sounds. All perceptual attributes were positively correlated, or negatively correlated, to the global loudness.

Also, global loudness equalization of sounds reduced the correlation between the perceptual attributes ratings and the sensory profiles of vehicles.

Additionally, based on the dispersion of the perceptual attributes ratings, it can be assumed that global loudness equalization forces listeners to more finely differentiate the perceptual attributes. It should then allow for a more detailed sensory profile differentiation.

Thirdly, a bootstrapping method was applied to simulate perceptual assessments of a large pool of virtual participants based on the actual assessments of a limited number of participants. As a result, the bootstrapping technique provides a better transcription of the information contained in the original assessments. Then, a principal component analysis was performed to reduce the dimensionality of data consisting of all perceptual attributes ratings. For sound samples not equalized in global loudness, the results indicated that 88% of the total variance in the initial data can be explained by a single dimension, which is the global loudness. For sound samples equalized in global loudness, the results showed that the sound profiles can be reduced to the first two principal dimensions, which retain about 69% of the total variance in the initial data. This suggests that discarding meaningless dimensions helps to de-noise data. Reflecting back on this study, it seems that a weak point of the methodology could be corrected for future studies. Indeed, for the vocabulary building sessions, all sound (with or without global loudness equalization) were used all together. Probably to be sure to have enough descriptors to describe both equalized and non-equalized case. Indeed, we later had the impression that more subtle descriptors might have been more useful to subtly describe the equalized sounds.

Future work should, in particular, explain objectively each of the perceptual attributes, the desire-to-buy and the significant principal component axes identified in this paper. For this purpose, in order to correlate subjective assessments and PC axes with objective physical or psychoacoustic metrics, it is planned to build linear regression models based on sparse modeling algorithms of multiple linear regressions (lasso and elastic-net) and nonlinear models based on random decision forest.

Acknowledgments

The authors wish to acknowledge the financial support from the "Natural Sciences and Engineering Research Council of Canada" (NSERC), "Bombardier Recreational Products" (BRP) and the "Centre de Technologies

Avancées BRP-UdeS” (CTA). Members of the dXBel project are also acknowledged for supporting the realization of this research work. The authors would like to thank Paul Massé from the marketing group of BRP for moderating the group discussions and for his support as well as subjects who participated to the subjective experiment described in this paper.

References

1. R.H. Lyon: An introduction to sound quality. *Journal of Sound and Vibration* 6 (2003) 6.
2. M. Gauthier: Identification et hiérarchisation des sources et chemins de transfert vibro-acoustiques d’un véhicule récréatif. Mémoire de maîtrise. Université de Sherbrooke, QC, Canada, 2017.
3. J. Blauert, U. Jekosch: Sound-quality evaluation—a multi-layered problem. *Acta Acustica United With Acustica* 83, 5 (1997) 747–753.
4. N. Otto, S. Amman, C. Eaton, S. Lake: Guidelines for jury evaluations of automotive sounds. *Journal of Sound and Vibration* (2001) 1–14.
5. S. Chen, D. Wang: Vehicle interior sound quality analysis by using grey relational analysis. *SAE International Journal of Passenger Cars – Mechanical Systems* 7, 1 (2014) 355–366.
6. T.G. Kim, S.-K. Lee, H.H. Lee: Characterization and quantification of luxury sound quality in premium-class passenger cars. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 223, 3 (2009) 343–353.
7. R. Guski: Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acustica United With Acustica* 83, 5 (1997) 765–774.
8. G. Lorho: Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction. Presented at the AES 119th Convention, New York, USA, 2005.
9. T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, S. Tervo: Concert hall acoustics assessment with individually elicited attributes. *Journal of the Acoustical Society of America* 194, 2 (2011) 835–849.
10. C.E. Osgood: The nature and measurement of meaning. *Psychological Bulletin* 49, 3 (1952) 197–237.
11. 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* 127 (2017) 270–283.
12. G. Von Bismarck: Timbre of steady sounds: Factorial investigation of its verbal attributes. *Acustica* 30 (1974) 146–159.
13. P. Susini, O. Houix, N. Misdariis, B. Smith, S. Langlois: Instruction’s effect on semantic scale ratings of interior car sounds. *Applied Acoustics* 70, 3 (2009) 389–403.
14. R.A. Kendall, E.C. Carterette: Verbal attributes of simultaneous wind instrument timbres. I – von Bismarck’s adjectives. *Music Perception* 4 (1992) 185–214.
15. T. Lageat, S. Czellar, G. Laurent: Engineering hedonic attributes to generate perceptions of luxury: Consumer perception of an everyday sound. *Marketing Letters* 14, 2 (2003) 97–109.
16. D. Dal Palù, B. Lerma, L.A. Grosso, L. Shtrepi, M. Gasparini, C. De Giorgi, A. Astolfi: Sensory evaluation of the sound of rolling office chairs: An exploratory study for sound design. *Applied Acoustics* 130 (2018) 195–203.
17. M.E. Altinsoy, U. Jekosch, The semantic space of vehicle sounds: Developing a semantic differential with regard to customer perception. *Journal of the Audio Engineering Society* 60, 1/2 (2012) 13–20.
18. T.H. Pedersen, N. Zacharov: How many psychoacoustic attributes are needed? *Proceedings of Euronoise 2008*, Paris, 2008.
19. J. Delarue, B. Lawlor, M. Rogeaux: *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research*. Elsevier, 2014.
20. H.T. Lawless, H. Heymann: *Sensory Evaluation of Food Principles and Practices*. Springer, NY, 2010.
21. P. Varela, G. Ares: Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International* 48, 2 (2012) 893–908.
22. K.A. Moussaoui, P. Varela: Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference* 21, 8 (2010) 1088–1099.
23. J. Delarue, J.-M. Sieffermann: Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference* 15, 4 (2004) 383–392.
24. H. Stone, J.L. Sidel: 6 – Descriptive Analysis, *Sensory Evaluation Practices*. Food Science and Technology, 3rd ed. Academic Press, 2004, pp. 201–245.
25. F. Bergeron, C. Astruc, A. Berry, P. Masson: Sound quality assessment of internal automotive road noise using sensory science. *Acta Acustica United With Acustica* 96, 3 (2010) 580–588.
26. M. Wankling, B. Fazenda, W.J. Davies: The assessment of low-frequency room acoustic parameters using descriptive analysis. *Journal of the Audio Engineering Society* 60, 5 (2012) 325–337.
27. N. Kaplanis, S. Bech, S. Tervo, J. Pätynen, T. Lokki, T. Waterschoot, S.H. Jensen: A rapid sensory analysis method for perceptual assessment of automotive audio. *Journal of the Audio Engineering Society* 65, 1/2 (2017) 130–146.
28. J. Berg, F. Rumsey: Identification of quality attributes of spatial audio by repertory grid technique. *Journal of the Audio Engineering Society* 54, 5 (2006) 365–379.
29. S. Le Bagousse, M. Paquier, C. Colomes: Categorization of sound attributes for audio quality assessment – a lexical study. *Journal of the Audio Engineering Society* 62, 11 (2014) 736–747.
30. V.-V. Mattila: Semantic analysis of speech quality in mobile communications: descriptive language development and mapping to acceptability. *Food Quality and Preference* 14, 5 (2003) 441–453.
31. A. Nykänen: *Methods for product sound design*. Doctoral dissertation, Luleå Tekniska Universitet, 2008.
32. G. Lemaitre, P. Susini: Timbre, sound quality, and sound design, in *Timbre: Acoustics, Perception, and Cognition* Siedenbuck K, Saitis C, McAdams S, Popper A, Fay R, Editors Springer Handbook of Auditory Research 69. 2019, pp. 245–272.
33. K. Genuit, A. Fiebig: *Sound Design of Electric Vehicles – Challenges and Risks*. InterNoise, Melbourne, 2014.
34. L. Langeveld, R. van Egmond, R. Jansen, E. Özcan: *Product sound design: Intentional and consequential sounds*, in *Advances in Industrial Design Engineering* Coelho DA, Editor IntechOpen. 2013.
35. P. Davies: Perception-based engineering: Integrating sound perception into engineering design. Presented at the InterNoise 47th I-INCE, Chicago, 2018.
36. E. Parizet, N. Hamzaoui, J. Jacquemoud: Noise assessment in a high-speed train. *Applied Acoustics* 63, 10 (2002) 1109–1124.

37. P. Susini, O. Houix, L. Seropian, G. Lemaitre: Is loudness part of a sound recognition process? *JASA* 146, 2 (2019) EL172–EL176.
38. S.-H. Shin, J.-G. Ih, T. Hashimoto, S. Hatano: Sound quality evaluation of the booming sensation for passenger cars. *Applied Acoustics* 70, 2 (2009) 309–320.
39. H. Fastl, E. Zwicker: *Psychoacoustics: Facts and Models*, 3rd ed. Springer, Berlin and Heidelberg 22, 2007.
40. G. Kwon, H. Jo, Y.J. Kang: Model of psychoacoustic sportiness for vehicle interior sound: Excluding loudness. *Applied Acoustics* 136 (2018) 16–25.
41. B. Efron: Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7, 1 (1979) 1–26.
42. B. Efron, R. Tibshirani: *An Introduction to the Bootstrap*. Chapman & Hall, 1993.
43. P. Hall: *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. New York, Springer-Verlag, 1992.
44. C.B. Barber, D.P. Dobkin, H.T. Huhdanpaa: The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* 22, 4 (1996) 469–483.
45. R. Palm: Utilisation du bootstrap pour les problèmes statistiques liés à l'estimation des paramètres. *BASE* 6, 3 (2002) 143–153.
46. I.T. Jolliffe: *Principal Component Analysis*, 2nd ed. Springer Series in Statistics. Springer-Verlag, New York, 2002.
47. W.J. Krzanowski: *Principles of Multivariate Analysis*. Clarendon Press, Oxford, 2000.
48. C. Duchesne: Multivariate image analysis in mineral processing, in *Advanced Control and Supervision of Mineral Processing Plants* Sbarbaro D, del Villar R, Editors. *Advances in Industrial Control*. London, Springer. 2010, pp. 85–142.
49. D.C. Montgomery: *Design and Analysis of Experiments*, 8th ed. John Wiley and Sons, 2013.
50. S.S. Shapiro, M.B. Wilk: An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965) 591–611.
51. M. Friedman: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *The Journal of the American Statistical Association* 32, 200 (1937) 675–701.
52. R. Broand, K.S. Age: Principal component analysis. *RSC Analytical Methods* 6, 9 (2014) 2812–2831.

Cite this article as: Benghanem A, Valentin O, Gauthier P-A & Berry A. 2021. Sound quality of side-by-side vehicles: Investigation of multidimensional sensory profiles and loudness equalization in an industrial context. *Acta Acustica*, 5, 7.