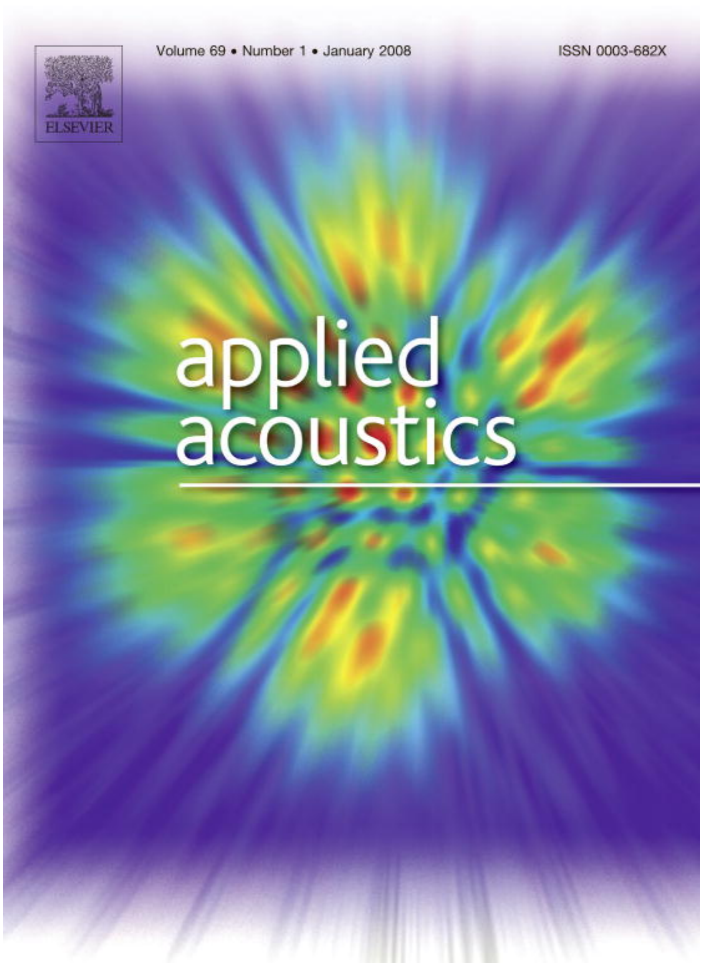

Volume 69 • Number 1 • January 2008
ISSN 0003-682X
ELSEVIER
applied acoustics

Available online at www.sciencedirect.com
ScienceDirect
Applied Acoustics 69 (2008) 12–22
applied acoustics
www.elsevier.com/locate/apacoust

# Analysis of car door closing sound quality

Etienne Parizet [a,*], Erald Guyader [a], Valery Nosulenko [b]

[a] *Laboratoire Vibrations Acoustique, Insa Lyon, 25bis Avenue Jean Capelle, 69621 Villeurbanne Cédex, France*
[b] *Institute of Psychology, Russian Academy of Sciences, 13, Yaroslavskaya Str., 129366 Moscow, Russia*

Received 4 November 2005; received in revised form 23 March 2006; accepted 11 September 2006

---

## Abstract

The perception of the noise coming from a car's door closure has been analyzed, the focus being put on the image of the quality of the car that the listener can have in mind while hearing the sound. Different experiments have been realized: a free sorting experiment for reducing the number of stimuli without any loss of generality, paired comparisons with similarity and preference ratings and, finally, free verbalizations analysis. The results have agreed on the importance of two timbre parameters, the frequency balance of the sound and its cleanness (only one temporal event should be audible). In particular, even if loudness had appeared as the most important sound feature in previously published studies, it did not in this one; the reason is probably that previous studies had focused on annoyance creating by sounds.

In a more general way, this study has proved the stability of the perceptual space derived from two different methods: a multi-dimensional analysis of similarity ratings and the analysis of free verbalizations.

---

## 1. Introduction

Noise from a vehicle door closing has two main functions. First of all, it indicates that the door was properly closed – which means that it has to be loud enough to be heard by the passenger. Moreover, it can contribute to the overall impression of the car; this is very important because closing the doors is one of the operations a customer can do while he is examining a car in the seller's hall. Kuwano et al. [1] have shown that cars with a pleasant sound (that descriptor was one from a group of descriptors used in a semantic differential experiment) were mainly thought as luxury ones. That study proved the strong link between the noise and the perceived quality of the whole car, in spite of the fact that the door is not a major component of the vehicle. The same study also checked that there does not seem to be any cultural influence on that perception, as ratings from Japanese and German listeners were very similar [2].

Many publications have dealt with door closing noise, but the methodology of listening experiments is never clearly explained and the sound features emphasized by these papers may be different. Loudness is claimed to be an important character by Fish and Franco-Jorge [3], Blommer et al. [4], Fridrich [5], Forbes and Wales [6], Champagne and Amman [7] or Petniunas et al. [8]. On the other hand, some other more complex parameters are suggested. They may be related to the frequency content of the sound (sharpness for Petniunas et al. [8], ratio between sound energy in the 1–3 kHz frequency band and energy in the 20–100 Hz frequency band for Malen and Scott [9], "predominant low frequency content" for Sellerbeck and Nettelbeck [10]). Temporal evolution of the signal was also emphasized by these latter authors, as well as by Hamilton [11] or Champagne and Amman [7].

In order to clarify all these points, it was decided to conduct a new study on that topic. One other goal of the car door seals supplier, which supported the study (Hutchinson–Paulstra), was to evaluate the influence of these door seals on noise perception.

---

[*] Corresponding author. Tel.: +33 4 72 43 81 21; fax: +33 4 72 43 87 12. *E-mail address:* etienne.parizet@insa-lyon.fr (E. Parizet).

0003-682X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2006.09.004

This paper will present the successive steps of the study. First of all, realistic noises were recorded from different cars in a controlled configuration. For some cars, the different seals were successively taken out of the door. As this led to a great number of stimuli, a first classification experiment reduced that number without losing any important aspect of the noises' context. Using a reduced number of stimuli, two other experiments were conducted in order to determine the perceptual space of such sounds. The first one used the verbalization analysis method and the other one consisted in evaluating similarities and preferences within pairs of sounds. From all the data thus obtained, important sound features could be determined.

## 2. Stimuli recordings

### 2.1. Door closing device

One important parameter for the closing sound of a car's door is the closure speed of that door. A special device was used for that purpose. The device was made up of a spring which could be compressed in a controlled way; that spring moved a beam which pushed the door. A photoelectric cell fixed on the door and on its frame measured the door's speed just before the closure. For each car, by trial-and-error, the minimum compression of the spring needed to close the door was determined; then it was decided to set the spring so that the final speed of the door was 25% higher than that minimum speed. Informal experiments showed that the initial energy thus transmitted to the door corresponded rather well to what an user does in order to be sure that the door is properly closed.

### 2.2. Recordings

Each car was entered in a semi-anechoic room. A dummy head (Bruel and Kjaer 4133) was installed outside the car, in a position corresponding to that of a driver closing his door (see Fig. 1). The driver's door was moved by the closing device described earlier and at least 4 recordings were realized for each car.

Fig. 1. Experimental set-up for sound recordings.

Sixteen cars were used; they were from eight different manufacturers and ranged from small cars to luxury cars.

Also, for two of these cars, door seals could be removed. According to the range of the car, up to three successive seal lines (one mounted on the frame and two on the door) can be used. Various combinations of seals removals provided four additional stimuli for one car and seven for the other one.

## 3. Experiment 1: reduction of the number of stimuli

The number of available stimuli (excluding the repetitions of each situation) was therefore 27 (16 cars plus 11 seals modifications on two cars). Such a number made it impossible to conduct a paired comparison experiment (the minimal number of pairs would have been $351 = \frac{27.26}{2}$, representing the upper half of a $27 \times 27$ matrix). Of course, a direct ranking of each noise would have been possible (using, for example, the equal interval method). But it has been observed that such a method is less accurate than paired comparisons [12]. Therefore, it was decided to reduce the number of stimuli, with the constraint that the important sound features should still be present in the reduced set of sounds; a classification experiment was used to reach that goal.

In everyday life, when exposed to a sound, listeners at first try to identify its source [13], which is a prerequisite to the evaluation of that sounds. This identification is realized through successive classifications (e.g. the sound is emitted by a car or a motorbike; if it is from a car, it can be a diesel engine or a gasoline one; eventually, depending on the listener's knowledge, the number of cylinders of the engine or even the brand of the car can give the basis for a next classification). In that way, classification is a very natural human activity. In the frame of listening test experiments, it can provide a tool for the evaluation of very different sounds (i.e., sounds emitted by very different sources), because for such stimuli context, other methods (e.g., multidimensional scaling ones) can fail as listeners' answers would be discontinuous.

In our study, sounds were not so different (at least, they were all identified as emitted from the same kind of sources), but the hypothesis was that a free sorting experiment could provide an useful tool to group sounds in clusters of stimuli having similar characteristics.

### 3.1. Procedure

One sample of each of the above-mentioned 27 situations was selected as stimulus. Also, for eight cars, a second sample was included in the data set, in order to check the repeatability of recordings. That gave a total amount of 35 sounds. All these sounds were presented through headphones (Sennheiser HD600) in a quiet room.

The whole experiment was conducted on a computer. At the beginning of the experiment, 35 buttons were presented to the listener (see Fig. 2 left). By clicking on each button, the listener could hear a sound which had been randomly assigned to the button. The task of the listener was to move all buttons in the upper part of the screen and to group them in families according to timbre similarities. He could freely distribute the buttons in the screen, build as many clusters as he felt necessary and was not asked to separate the families in a meaningful way (i.e., the distance between different families should not represent any perceptual distance). In the example presented in Fig. 2 right, the listener had made five families.

Thirty-one listeners took part in that experiment. They were members of the laboratory or students.

### 3.2. Results

It appeared that this experiment was not too difficult nor too long. A typical duration was a little less than half an hour, though listeners did not hesitate to play sounds many times. The average number of played sounds was 610 (minimum: 258, maximum: 1427), which was possible because such sounds were quite short (less than 2 s).

The number of categories created by listeners varied between 4 and 8 for most of them, though one subject defined 18 groups (see Fig. 3). From all individual results, a $35 \times 35$ matrix $M$ was computed, in which

$$M(i,j) = 1 - N(i,j) \tag{1}$$

where $N(i,j)$ is the proportion of listeners who grouped the sounds labelled $i$ and $j$ in the same category.

$M$ was considered as a distance matrix (indeed, it was not a distance matrix; in particular, the distance criteria $M(i,j) + M(j,k) \geqslant M(i,k)$ was not fulfilled). A hierarchical cluster analysis [14] was computed from that matrix and the dendrogram thus obtained is presented in Fig. 4.

It appeared that this dendrogram could be cut in six parts, labelled Groups 1–6 in Fig. 4. The computation of adjusted Rand Index [15,16] using the bootstrap technique showed that reliable clustering of the set of sounds had 5 to 7 groups and the 6-levels partition was selected.

It was noticed that, when two recordings from the same car were present, they belonged to the same group; these repeated recordings are represented by the pairs $(2,3)$, $(7,8)$, $(9,10)$, $(11,12)$, $(20,21)$, $(25,26)$, $(29,30)$ and $(31,32)$. This confirmed the repeatability of the recording set-up.

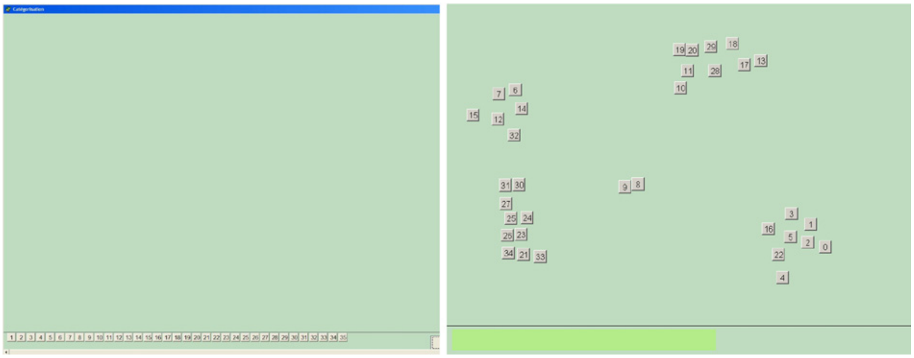


Fig. 2. Computer's screen of the classification experiment. Left: beginning of the experiment; right: end of the experiment, for a given listener.

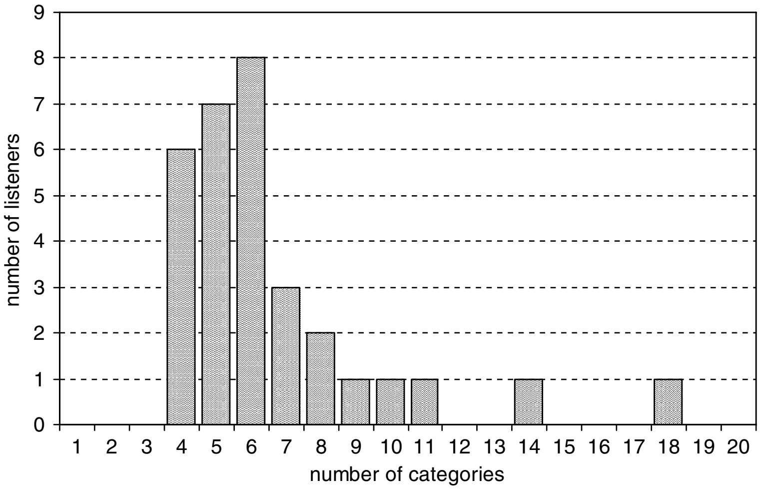


Fig. 3. Histogram of the number of categories.

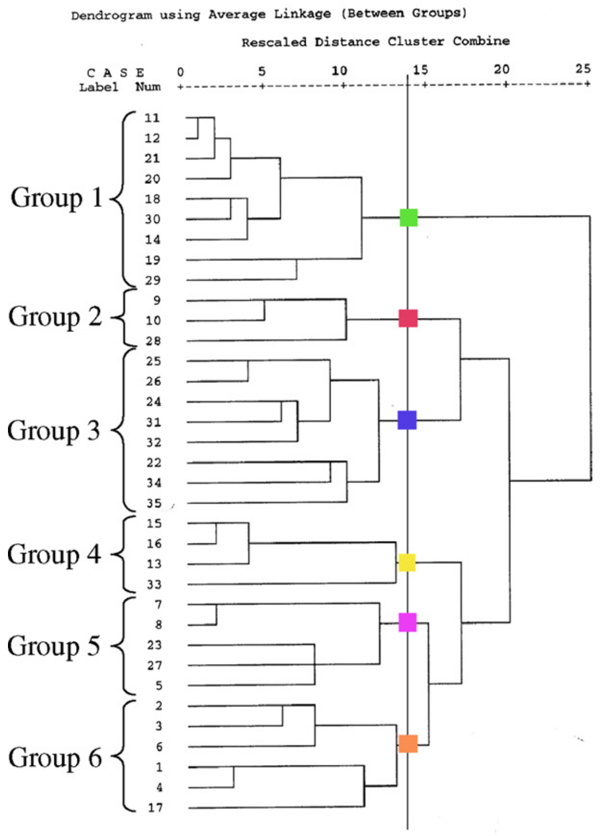


Fig. 4. Dendrogram computed from the classification experiment results.

On the other hand, when seals were removed from the car, sounds could jump from one group to another. Such sounds are labelled (11, 13, 14, 15, 16, 17, 18, 19) for one car and (2, 4, 5, 6) for the other one. Therefore, as could be expected, seals have a great influence on the sound image of the door.

## 4. Experiment 2: determination of the perceptual space

### 4.1. Procedure

Within each group of similar sounds, two representative sounds were selected on the basis that they were the closest ones from all sounds of their group (in the meaning of the pseudo-distance defined by the classification matrix *M*). They had the following labels (see Fig. 4):

- group 1: 30 and 18,
- group 2: 9 and 28,
- group 3: 26 and 32,
- group 4: 13 and 16,
- group 5: 8 and 27, and
- group 6: 3 and 6.

These 12 sounds were supposed to represent the whole context of recorded sounds. They were used in a paired comparison listening test. After a preliminary presentation of these sounds, each possible pair was presented in a random order (through headphones). After listening to a given pair (as often as he felt necessary), the subject had to evaluate the similarity between the two sounds, by moving a cursor on a continuous scale, labelled at each side (from "extremely similar" to "extremely different"). Then he had to listen to the pair again and to answer to the following question: "Which is the sound evoking the best quality of the door for you?". There were three possible answers, as the listener could select one of the sounds or not (the "equal" answer was allowed, because that procedure had some advantage on the forced-choice one [17]).

Forty people took part in this experiment, the jury being balanced in two ways (sex and age). There were 19 women and 21 men and, in each sex group, half of subjects were between 30 and 45 and half between 46 and 60. They did not belong to the laboratory and were paid for there participation (10 Euros).

### 4.2. Results

#### 4.2.1. Comparison of the door quality evoked by sounds

The average preference probabilities were analyzed using a linear computation. For each sound, a score $S_i$ was computed by:

$$S_i = \frac{1}{N} \sum_{j \neq i} P_{ji} \tag{2}$$

where ($P_{ij}$) are the preference probabilities within pairs (i.e., $P_{ij}$ is the probability with which the sound *j* was preferred to sound *i*) and *N* is the number of listeners.

Such a model proved to be valid, as estimates of preference probabilities (obtained by $\widetilde{P}_{ij} = S_j - S_i$) were closed to the measured ones $P_{ij}$ (the correlation coefficient between the set of measured and estimated probabilities was greater than 0.94).

This analysis was first conducted for the two subdivisions of the jury. It appeared that there was no difference between male and female judgments (Fig. 5, left diagram. On that figure and the following one, it should be noted that a **high value** of the merit score indicates a **poor quality** of the door evoked by the sound; in the following, these values will be named *demerit scores*). On the other hand, results computed for younger and older people showed some discrepancy (Fig. 5, right diagram); the difference was statistically significant ($p < 0.05$) for sounds 28 and 6. The hierarchy between sounds was similar, but the maximum values were greater for younger subjects, because younger subjects expressed clearer preferences within pairs, which gave preference probabilities closer to their extreme values (0 or 1). In the preference probabilities averaged over each of these two groups of listeners, the value $L_{75}$–$L_{25}$ (difference between the third and the first quartiles) was equal to 0.30 for older subjects and 0.55 for younger ones.

Demerit scores computed over the whole panel of listeners are shown in Fig. 6. In that figure the average values and

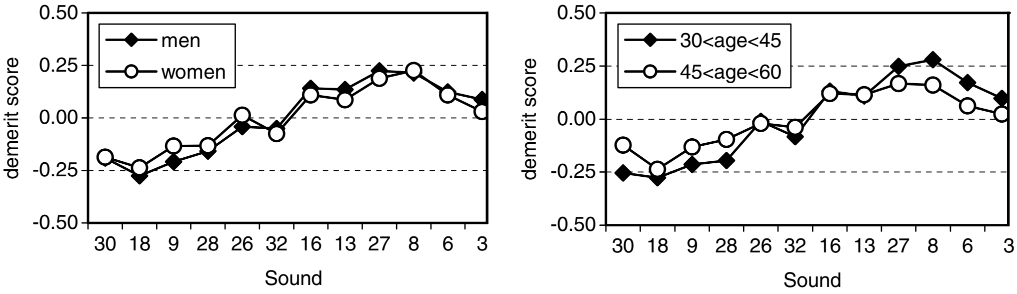


Fig. 5. Merit scores of sounds. Left: results for women and men. Right: results for younger and older subjects.

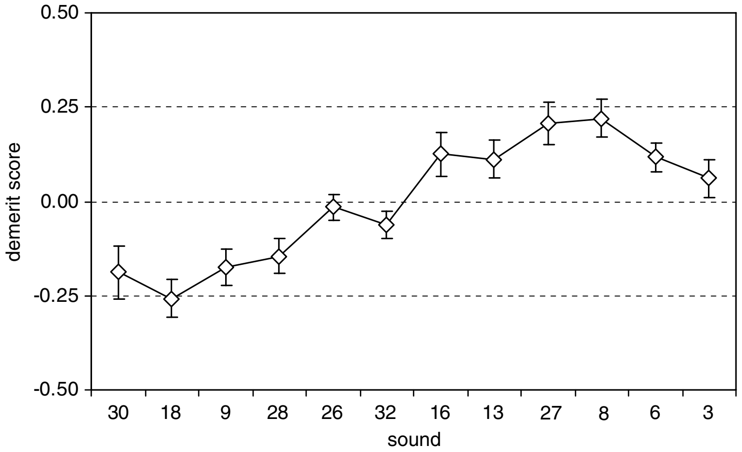


Fig. 6. Demerit scores computed over the whole panel.

the confidence interval of scores ($p = 0.05$) are represented. The confidence intervals could be easily computed because the linear computation of scores (see Eq. (2)) could be realized for each listener. Sounds from each group have similar scores, which confirmed that the evaluation was made on the basis of sound timbre. With regard to the range of cars, there was no clear link between the price of the car and the sound quality of its driver's door. For example, sounds 27 and 8 in Fig. 6, which appeared as the worst ones, were recorded on cars which were more expensive than sound 26, in which sound quality was better.

### 4.2.2. Similarity between sounds

All individual results were converted in numbers from 0 (corresponding to "sounds are extremely similar") to 1 ("extremely different"). These numbers were averaged over the 40 listeners and a hierarchical cluster analysis was computed (Fig. 7).

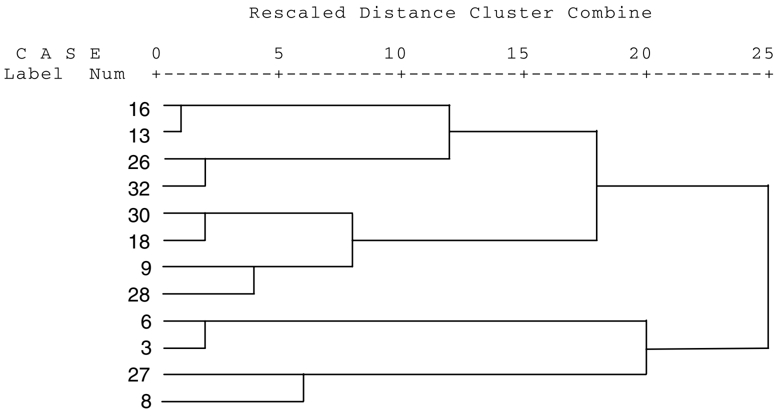


Fig. 7. Clustering of sounds, obtained from similarity ratings.

These results confirmed those obtained from the free sorting experiment: the two sounds representing each group (as determined in the first experiment) were still perceived as close to each other in the second listening test. Moreover, the set of data is still organized in 6 groups. Of course, the comparison between Figs. 4 and 7 shows some discrepancies: for example, the classification experiment indicated that the second group was closer to the third one than to the first one, which does not appear in Fig. 7. This is certainly due to the limit of the free sorting experiment: listeners were asked to group sounds according to their timbre. But they did not have to give any information about the perceptual distance between groups: that information is therefore missing.

### *4.2.3. Perceptual space*

The perceptual space was determined from an Indscal analysis [18] of similarity results. Without going too much into details, this analysis provides a set of sound coordinates $x_{i,s}$ ($i$ denoting sound) over each axis of the perceptual space ($s$ denoting the axis) and a set of individual weightings $w_s^t$, $t$ denoting the listener, such that the individual similarities $\delta_{ij}^t$ (between sounds $i$ and $j$) are approximated by $d_{ij}^t = \sqrt{\sum_{s=1}^{S} w_s^t \cdot (x_{i,s} - x_{j,s})^2}$, $S$ being the number of dimensions selected. In that case, choosing $S = 3$ (the curve relating Kruskal's stress to the number of dimensions showed an elbow for that value) provided a good approximation of the set of measured individual similarities ($\delta_{ij}^t$) and Fig. 8 represents the positions of sounds in that three-axis perceptual space.

By listening to sounds, the two first axes could be interpreted:

- The first axis was related to the frequency content of sounds. Sounds 9,18, 28 and 30, which are located on the right-hand side of axis 1, contained less energy in the high frequencies than sounds 6 or 27.
- The second axis was determined by the cleanness of sounds. In sound 30, only one impulse could be heard, while several ones could be detected in sounds 13 or 16.

Such various timbres can be easily detected on a time-frequency analysis of signals. As an example, Fig. 9 shows a wavelet analysis (realized by the 01 dB-Metravib *dBSonic* software) of the left channel of sounds 30, 16 and 6. The difference in frequency balance between sounds 30 and 6 can be noticed, as sound 6 contained more energy above 500 Hz. Also, three separated events are clearly visible on the analysis of sound 16.

These timbre aspects could be represented by two metrics:

- For the first axis, the sharpness as defined by Aures [19] or the spectral centroid [20]; the correlation coefficient between the values of these metrics and the sound coordinates on the axis was −0.90 for sharpness and −0.93 for spectral centroid.
- For the second axis, an indicator was derived from the temporal loudness calculation [21]. The algorithm proposed by Zwicker to take temporal integration and temporal masking into account was used to compute the

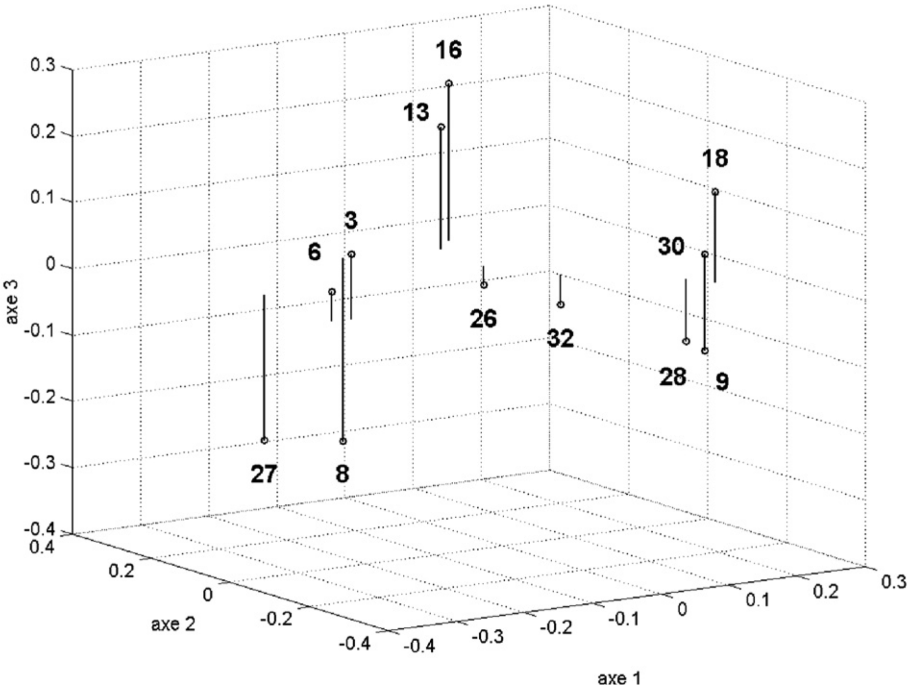
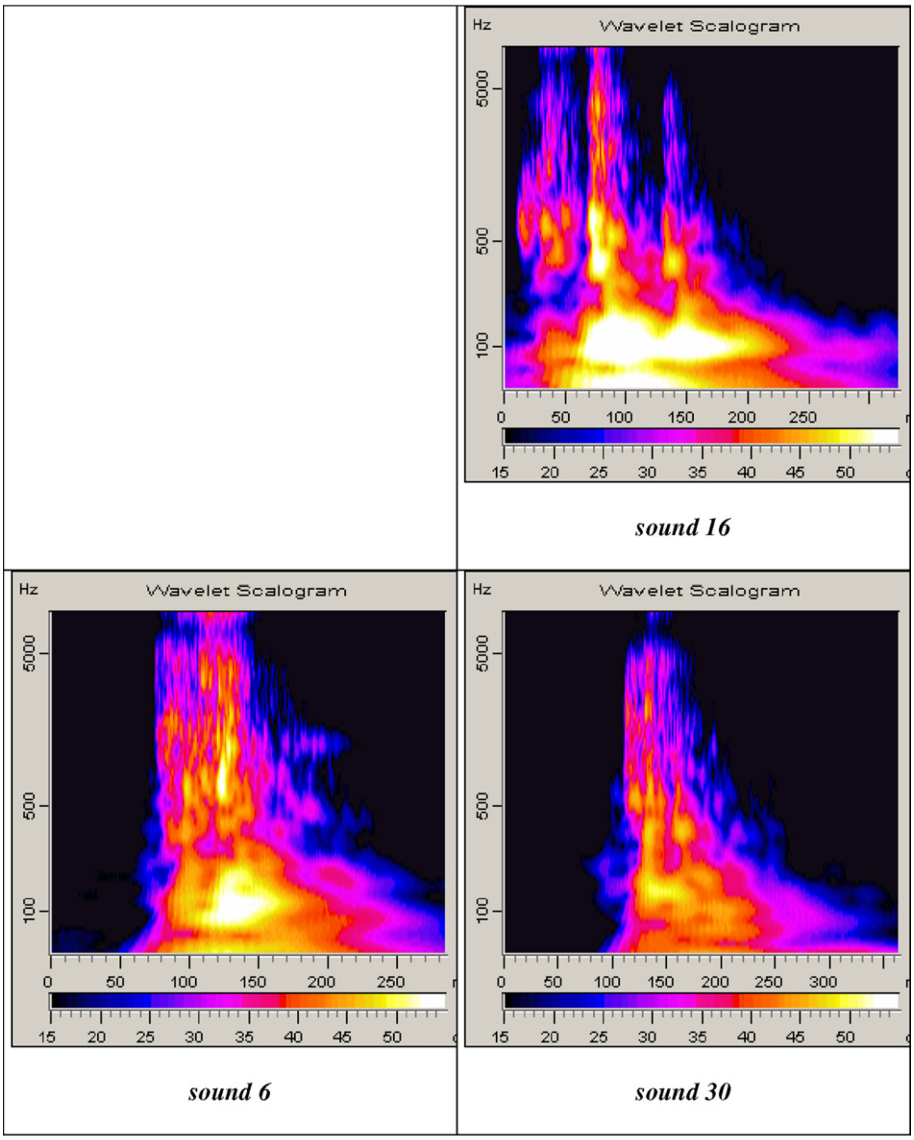


Fig. 8. Perceptual space of door closing sounds.



**sound 16**

**sound 6**

**sound 30**

Fig. 9. Wavelet analysis of sounds labelled 30, 16 and 6 (left channel). The drawings have been placed according to the relative positions of theses sounds on the (1–2) plane of the perceptual space.

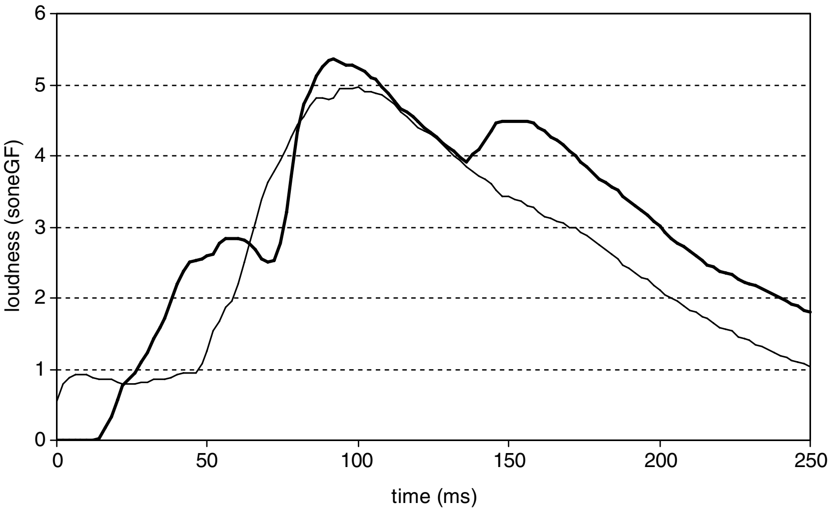


Fig. 10. Instantaneous loudness, computed according to Zwicker's procedure, of two sounds. (——): sound 16; ——: sound 9.

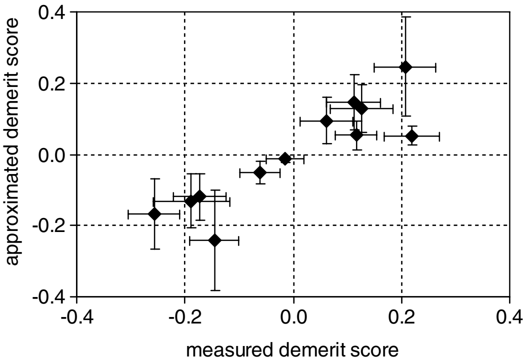


Fig. 11. Comparison of predicted and measured scores.

instantaneous loudness; two examples are shown in Fig. 10 (sounds 9 and 16). Three different events can clearly be detected for sound 16. The proposed indicator was not based on the maximum value of loudness (on the presented examples, this value is very similar for the two sounds), but on the temporal evolution of the curve. The correlation coefficient between the values thus obtained and the coordinates of sounds on the second axis was 0.87.

A demerit score model was computed from the spectral centroid (hereafter designed as $X_1$) and the indicator describing the second axis of the perceptual space ($X_2$). Linear scores could be correctly approximated ($R^2_{\text{adj}} = 0.76$, $F(2,9) = 18.2$, $p < 0.01$) by a linear equation involving $X_1$ and $X_2$:

$$L_i = C + \alpha X_1 + \beta X_2 \tag{3}$$

The comparison between measured and predicted demerit scores shows a correct agreement (Fig. 11), apart for sound 8, for which the predicted score (0.05) is significantly smaller than the real one (0.22). That difference may be due to an additional sound feature, corresponding to the third dimension of the perceptual space which could not be interpreted (in Fig. 8, it can be seen that sound 8 has an important coordinates on the third axis).

## 5. Experiment 3: verbalizations analysis

### 5.1. Procedure

It was decided to get some knowledge of the perceptual space using another method: the verbalizations analysis. That method [22] assumes that, in a comparison task, there is a strong relation between the cognitive processes involved in performing the task and the processes identified in verbal reports produced during or after it. The experimental method is also based on paired comparisons; in our study, the task of the listener was, first of all, to evaluate the similarity between the two sounds and, then, to select the one evoking the best quality of the door. Finally, the listener had to freely describe the similarities and differences he could hear between stimuli with regard to that criteria (quality of the door) and to justify his preference choice. All his verbalizations were recorded on a two-channel recorder, on the second track on which one of the audio channels was simultaneously recorded.

As such an experiment and the corresponding analysis are very time-consuming, it was decided to use only six sounds; they were the best representative sounds of each group and were included in the set of 12 sounds used in the previously described experiment. Namely, these sounds had the numbers 30, 9, 26, 13, 8 and 3 (see Fig. 4). Also, only 11 subjects participated to that experiment (6 women and 5 men); they were not used to listening tests (in particular, they had not participated to one of the previous experiments) and were paid for their participation (10 Euros).

### 5.2. Results

The analysis of such verbalizations has already been exposed in details for musical sounds [22] or industrial ones [23]; briefly, it consists of detecting the “verbal units” (i.e., successive parts of the verbalizations referring to separate characteristics of sounds). The number of such verbal units varied between 84 and 178, the average being 120 (as 15 pairs were submitted to listeners, that number represents an average of eight verbal units for each pair). All these excerpts are then analyzed in different steps: for example, did the verbal unit refer to a similarity or a difference between sounds? Was it expressed on a general basis (for example, “these stimuli are very different”) or on a concrete one (“the second sound is louder”)? The last step consists of understanding the meaning of the feature mentioned by the listener and to group together similar features (for example, “loud”, “high level”, and so on).

In a first step, demerit scores were computed, using Eq. (2) in which $N = 6$. For each sound, the score was close to the one obtained in the previous experiment (Fig. 12), which indicated a good agreement of the two panels about what such a sound should be.

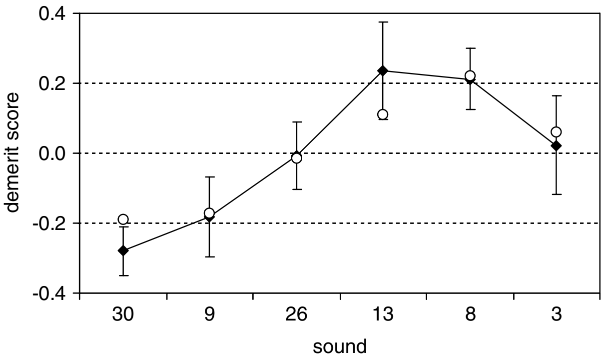


Fig. 12. Demerit scores for the 11-listeners jury of the verbalizations experiment (closed symbols). Values obtained from the 40-listeners jury of the paired-comparison experiment are shown as open symbols for comparison.

The number of times each descriptor was expressed by listeners (number of citations) was evaluated and nine main descriptors could be identified. They were labelled as "sharp", "pleasant", "loud", "accurate" ("the sound is clear and precise"), "high-quality car" ("it sounds like a high-quality car"), "damped", "well closed", "quick", "small car", "annoying", and "secure". The sum of the citations of theses descriptors explained 98% of the whole number of citations. The relative uses of these descriptors are shown in Fig. 13; it can be seen that "sharp", "pleasant", "loud", "accurate", and "high-quality car" were the most often cited descriptors. On the other hand, the "annoying" category was rarely used, which could be expected because listeners were oriented on the evaluation of the quality of the car and not on any annoying descriptor of sounds.

In the database derived from the analysis of verbalizations, each descriptor was given a positive or negative label (for example, a sound could be described as "loud" or "not loud"); it was then possible, by using a procedure described in [23], to draw the "verbal portraits" of sounds, indicating their main features. That method makes it possible to establish significant characteristics that determine estimation and preference in human judgments as well as the "weight" of each of them.

The examples for sounds 30, 9, 13 and 8 are shown in Fig. 14. In that figure it can be seen that sounds 9 and

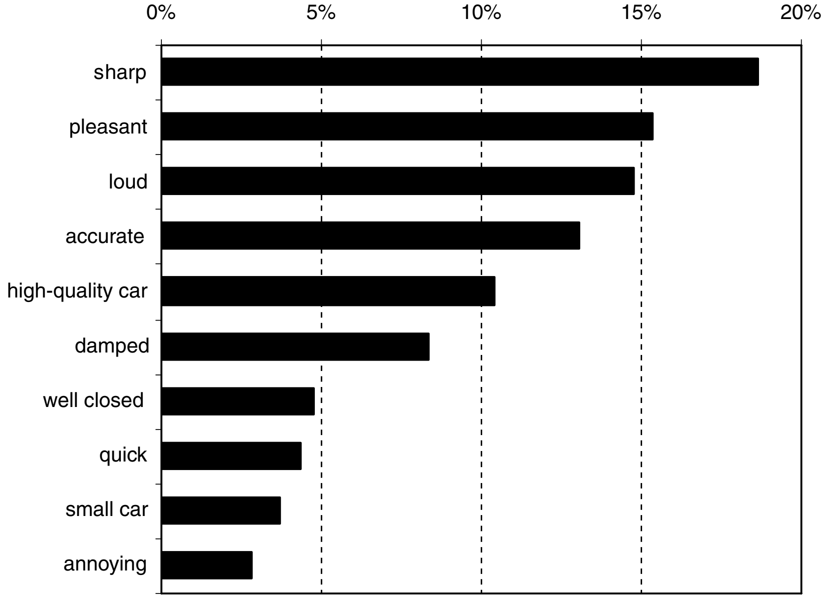


Fig. 13. Occurrence rates of descriptors families.

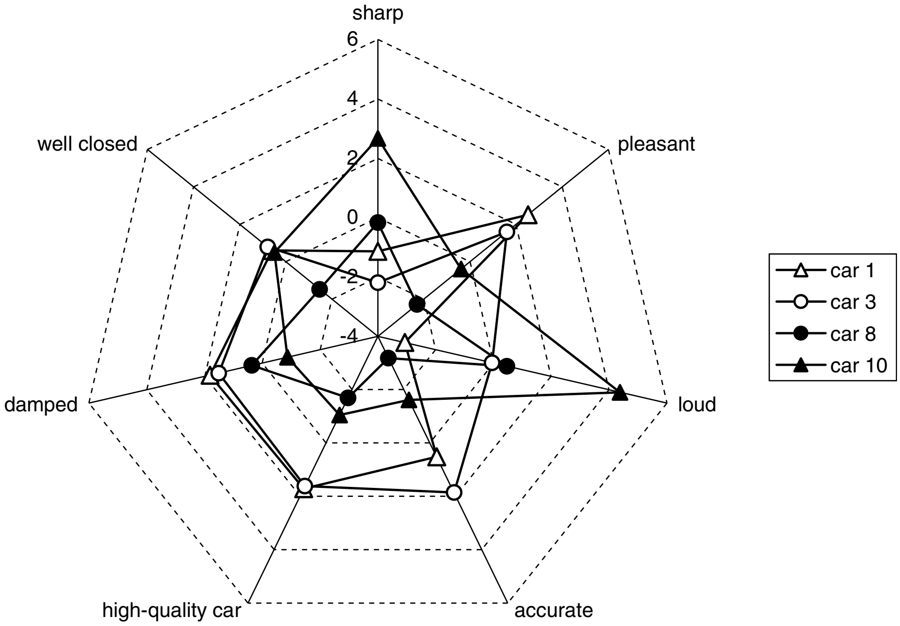


Fig. 14. Verbal portraits of four sounds.

13, though described as equally loud, evoked a very different quality of the cars. Sounds had very different loudness (the $N5$ values [21] varied from 3.5 to 8 SoneGF) and that loudness was clearly noticed by listeners, as can be seen in Fig. 13 (almost 15% of verbalizations were related to that feature); but it did not contribute so much to sound quality.

On the other hand, Fig. 14 suggests that descriptors "sharp", "damped" and "accurate" could be more closely related to sound quality. In order to make clear the relation between verbal portraits and the perceptual space obtained from the similarity ratings, it was decided to compute the correlation coefficients between the values attributed to sounds according to each descriptor (from the verbalizations analysis) and sounds' coordinates on the axis of the perceptual space shown in Fig. 8. It appeared that the first axis was correlated with the "sharp" ($R = -0.94$) and the "damped" ($R = 0.95$) aspects and the second one was correlated with the "accurate" descriptor ($R = -0.90$), which confirmed the relevance of sound features obtained from the previous experiment.

These relations confirmed the conclusions derived from the direct interpretation of the perceptual space. The two experiments, based on very different methodologies, thus gave similar results, indicating a great stability of cognitive processes used by listeners during the different tasks.

## 6. Discussion and conclusion

As mentioned by Kuwano et al. [2], listeners have common expectations about what sound quality of a car door closing should be. The study conducted by Kuwano et al. involved German and Japanese listeners while that study was conducted with French listeners only (with a greater age range than the previous one): though the cultural and age differences (which can be important for some other sound quality applications), results of both study are similar.

The fact that loudness was not a major parameter of sound quality, contrary to what was related by previous studies ([3–8]), can be explained by the differences between the questions asked to subjects. In the cited studies, listeners had to evaluate the annoyance of sounds, as if they were in their living room and could hear their neighbour coming back home and closing his car's door; on the contrary, in our study, listeners were in the position of a customer closing his own car's door and evaluating the quality of his car from the sound. In that case, annoyance cannot be a relevant descriptor; as can be seen in Fig. 13, annoyance was not often mentioned by listeners, which confirmed that they really turned their attention to the task they were asked to achieve.

Sound parameters determined from these experiments are in accordance with results from the literature: the sharp/damped descriptor appeared in the study from Malen and Scott [9] or Sellerbeck and Nettelbeck [10], and the fact that the audibility of different events decrease sound quality is consistent with the temporal evolution cited by Hamilton [11] or Champagne and Amman [7], though there are not so many details about it in these papers.

The demerit score model proposed in that study can now be used by the supplier to evaluate the influence of each seal in the sound perception (as mentioned in part 3.2, removing a seal can have a strong influence on sound quality).

## Acknowledgements

The authors want to acknowledge Hutchinson–Paulstra, which supported the study.

## References

[1] Kuwano S, Fastl H, Namba S, Nakamura S, Uchida H. Subjective evaluation of car door sound. In: Proceedings of the Sound Quality Symposium SQS; 2002.

[2] Kuwano S, Fastl H, Namba S, Nakamura S, Uchida H. Quality of door sounds of passenger cars. In: Proceedings of the ICA; 2004. p. 1365–68.

[3] Fish D, Franco-Jorge M. Towards generic design criteria for vehicle door shut exterior noise quality. In: Proceedings of the Internoise 97; 1997.

[4] Blommer M, Otto N, Wakefield G, Feng B, Jones C. Calculating the loudness of impulsive sounds. Publication of the Society of Automotive Engineers SAE 951311; 1995.

[5] Fridrich R. Investigating calculated loudness ISO532 for evaluating impulsive sounds. Publication of the Society of Automotive Engineers SAE 911088; 1991.

[6] Forbes J, Wales R. Test process for best-in-class door closing sound quality. In: Proceedings of the IBEC '95, Body Design and Engineering; 1995. p. 145–51.

[7] Champagne A, Amman S. Vehicle closure sound quality. Publication of the Society of Automotive Engineers SAE 951370; 1995.

[8] Petniunas A, Otto N, Amman S, Simpson R. Door system design for improved closure sound quality. Publication of the Society of Automotive Engineers SAE 1999-01-1681; 1999.

[9] Malen D, Scott R. Improving automobile door-closing sound for customer preference. Noise Control Eng J 1993;41(1):261–71.

[10] Sellerbeck P, Nettelbeck C. Door operating sound improvement based on jury testing and system analysis. In: Proceedings of the CFA/DAGA '04, 977-978; 2004.

[11] Hamilton D. Sound quality of impulsive noises: an applied study of automobile door closing sounds. Publication of the Society of Automotive Engineers SAE 1999-01-1684; 1999.

[12] Parizet E, Hamzaoui N, Sabatié G. Comparison of listening test experiments: a case study. Acta Acust United Acust 2005;91: 356–64.

[13] Rosenblum L. Perceiving articulatory events. In: Neuhoff J, editor. Ecological psychoacoustics. Elsevier Academic Press; 2004.

[14] Jain A, Murty MN, Flynn PJ. Data Clustering: a review. ACM Comput Surv 1999;31(3):264–323.

[15] Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 1971;66:846–50.

[16] Hubert L, Arabie P. Comparing partitions. J Class 1985:193–218.

[17] Parizet E. Paired comparison listening tests and circular error rates. Acta Acust United Acust 2002;88:594–8.

[18] Carroll J, Chang J. Analysis of individual differences in multidimensional scaling via an $n$-way generalization of Eckart–Young decomposition. Psychometrica 1970;35:283–319.

[19] Aures W. Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen. Acustica 1985;58(5):282–90.

[20] Grey JM, Gordon JW. Perceptual effects of spectral modifications on musical timbres. J Acoust Soc Am 1978;63:1493–500.

[21] Zwicker E. Procedure for calculating loudness of temporally variable sounds. J Acoust Soc Am 1977;62:675–81.

[22] Samoylenko E, Nosulenko V, McAdams S. Systematic analysis of verbalizations produced in comparing musical timbres. Int J Psychol 1996;31(6):255–78.

[23] Nosulenko V, Parizet E, Samoylenko E. La méthode d'analyse des verbalisations libres: une application à la caractérisation des bruits de véhicules. Soc Sci Inform 1998(374):593–611.