Car door closure sounds: characterization of perceptual properties through analysis-synthesis approach

Marie-Céline Bezat, Vincent Roussarie, Thierry Voinier, Richard Kronland-Martinet, Sølvi Ystad

19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

CAR DOOR CLOSURE SOUNDS: CHARACTERIZATION OF PERCEPTUAL PROPERTIES THROUGH ANALYSIS-SYNTHESIS APPROACH

PACS: 43.66Jh

Bézat, Marie-Céline[1,2]; Roussarie, Vincent[2]; Voinier, Thierry[1]; Kronland-Martinet, Richard[1]; Ystad Sølvi[1]
[1]CNRS-Laboratoire de Mécanique et d'Acoustique; 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France; {voinier, kronland, ystad} @lma.cnrs-mrs.fr
[2]PSA PEUGEOT CITROEN, 2 route de Gisy, 78943 Velizy Villacoublay, France; {marieceline.bezat, vincent.roussarie} @mpsa.com

ABSTRACT
The aim of this study is to identify perceptually pertinent parameters for the evaluation of car door closure sounds. For this purpose, perceptual properties of recorded sounds were first evaluated by sensory metrology. Then, an analysis-synthesis approach was chosen in order to identify perceptually pertinent signal parameters. The analysis part of this process first consisted in decomposing the sound in several independent impact sources using the Empirical Modal Decomposition method. Each impact is then modeled by a set of gains and damping factors in each critical band (ERB). These parameters were further used to synthesize sounds related to various aspects of the door closure sound with a real-time tool. This approach allowed for the generation of realistic, synthesized car door closure sounds that preserve perceptual properties with a reduced number of signal parameters. Listening tests finally allowed for the observation of the influence of the main signal parameters on the perceptual properties of such action-related impact sounds.

INTRODUCTION
Door closure sounds are of interest for impact sound research. From an acoustical point of view, such sounds are complex and composed by several impacts, such as the latch mechanism and the door/panel impact. From a perceptual point of view, it conveys complex perceptual properties: it tells whether the door is well closed and contributes to the overall impression of the car quality. The relations between sounds and perception have already been studied, but are still not well understood. A recent study [1] proposes timbre parameters as frequency balance and cleanness (only one temporal event is required) to predict preference ratings. This model is based on listening tests by paired comparison for similarity and quality ratings of 12 sounds and statistical links between signal parameters and perceptual dimensions. It appears, however, that this simple model is not sufficient to unambiguously predict perceptual properties. As shown for instrumental sounds [2], this issue could be addressed with synthetic sounds directly controlled with signal parameters, and especially designed to cover the perceptual space: on the one hand, we directly observe the effect of each signal parameter on perceptual properties in order to identify the true underlying parameters; on the other hand we mix the problem of finite number of data with a good repartition of stimuli in the perceptual space. Moreover we propose the sensory metrology instead of similarity rating to access perceptual properties. This methodology consists in extracting salient descriptors that characterize the sound space without any assumption on the continuity and orthogonally of dimensions. On the one hand, sensory descriptors that describe sounds *per se* are easier to link to signal parameters than impressions or events; on the other hand, they contain sufficient information to predict sound quality [3].
We thus propose to define an analysis synthesis model based on perceptual analysis of real sounds. The model has to reproduce the sound from a perceptual point of view and to be controlled by few parameters to allow an easy observation of the influence of signal parameters on perceptual properties.

PERCEPTUAL ANALYSIS OF REAL SOUNDS

We first conducted a sensory profile on real door closure sounds and pointed out the relevant perceptual properties to investigate.

10 subjects (audiologically controlled) (average of 45 years old, all women) effectuated the sensory test of 26 outdoor closure sounds, measured in a Semi Anechoic Room from vehicles of different brands and segments. The stimuli were chosen for their variety from a previous classification experiment with a large number of data. 8 sessions of 2 hours were organized and lasted for 2 months. The sensory profile was elaborated threw 4 stages:

1. A semantic learning stage: the subjects had to establish a lexicon of descriptor words which illustrated at best the sounds *per se*. No information regarding the source or the feeling was given (2 sessions)
2. A sensory learning stage: the subjects were trained to discriminate the sounds with the proposed descriptors, (2 sessions)
3. A metric learning stage: the subjects were trained to use a descriptor scale (0-10), (2 sessions)
4. An evaluation stage: the subjects evaluated all the sounds three times. (2 sessions)

The panel performances were controlled for every descriptor by 3 criteria: discriminability of stimuli, repeatability of subjects, and consensus between subjects for stimuli evaluations. Discriminability and repeatability were verified with variance analyses (ANOVA) with 3 factors (product, subject, and repetition). Consensus was verified with the percentage of inertia carried by the first axis of the PCA (Principal Component Analysis) {Product, Judge} by descriptors. It estimated the dispersion of the judges on each component. With inertia on the 1st axis higher than 70%, the consensus was satisfying for all descriptors.

Table I.- ANOVA with 3 factors (product, subject, and repetition)

	ddl	INTENSE	BON'M	ESPACHOC	KE	VIBRANT	Comments
Product 1	25	36,23***	29,01***	14,09***	20,77***	13,4***	Discriminability
Judge 2	8	13,48***	16,65***	9,84***	24,41***	15,38***	Scale effect
Repetition 3	2	1.31	0.47	3.36	1.09	1.94	Repeatability
Interaction 1-2	200	1,3*	2,76***	3,99***	2,97***	3,75***	Interindividual differencies
Interaction 1-3	50	0,91	0,98	0,85	0,95	0,80	Repeatability
Interaction 2-3	16	7,23***	1	1,99*	2,92***	1.09	lack of repeatability of one or more judges

Table II.- Descriptors of the sensory profile

Descriptor	Translation	Definition
INTENSE	LOUDNESS	Which has an high loudness
BON'M	*BON'M*	Which has a strong ''BON'M'' sound
KE	*K*	Which has a strong ''KE'' sound
ESPACHOC	INTER CLICK TIME	Which has a strong duration between 2 impacts
VIBRANT	VIBRATION	Which has strong oscillations

Some of the descriptors are quite classical (LOUDNESS), others are onomatopoetic ones (KE, BOMN). With the sound imitations, it is indeed more straightforward to find underlying signal parameters. Some of the descriptors indicate possible defaults for a few sounds of our corpus (VIBRATION, ESPACHOC), but are not relevant for all sounds. Sounds with a strong KE are described by automotive experts as sounds with a strong participation of latch mechanism and sounds with a strong BOMN are described as sounds with a weak participation of latch mechanism and a strong low frequency resonance of the door panel.

We focus for the analysis-synthesis model elaboration on the 3 main descriptors: LOUDNESS, BOMN, and KE. Those descriptors are relevant for characterizing all door closure sounds and the descriptors BOMN and KE are related with main sources (latch mechanism and door panel). The descriptor LOUDNESS appeared as we chose to study doors closed with more or less energy. We aimed at recreating sounds from those 26 door closure sounds that:

- are recognized as door closure sounds,

- are perceptually close to real sounds, and that in particular contain as many perceptual properties as real sounds (LOUDNESS, BONM and KE).

ANALYSIS-SYNTHESIS MODEL

Physically-based impact models are used for psychomechanical research [4], [5], but door closure sounds resulting from several complex impact sounds are difficult to model. Other signal-based models such as granular synthesis could perfectly reproduce door closure sounds, but those models need a large number of parameters that are not directly related to perceptual properties. We here propose a ''perceptual'' model based on our previous perceptual analysis and controlled by few parameters to allow an easy exploration of the door closure sound space. The model described below is derived from an analysis-synthesis impact model developed by Aramaki [6], [7].

Decomposition in elementary sounds using the Empirical Modal Decomposition method

The door closure sounds are first decomposed in elementary sounds: the latch mechanism part and the low frequency resonance part. As the latch part presents impacts with broadband energy, a simple filter is not sufficient to separate it from the resonance part. Hence, we used the Empirical Mode Decomposition method [8], [9]. Its principle consists in repeatedly identifying zero-mean AM- FM modes by locally separating a 'fast' contribution (corresponding to the latch part) from a ‘slower’ tendency (corresponding to the resonance part). The initial door closure sound is perfectly reconstructed by addition of the two parts.

Analysis

The two elementary sounds are then analysed with a time-scale ERB representation from which gains and damping factors in each critical band (ERB) are extract for the synthesis.
The time-scale ERB representation is obtained by decomposing the signal with a filterbank composed of Gaussian functions which center frequencies correspond to the ERB scale (integer values of B). The bandwidth of each filter is designed to minimize information loss. The ERB-scale function is given by the relation [10]:

$$ERBnumber = 21.4\log(4.37F + 1) \qquad \text{(Eq. 1)}$$

where F is the frequency in kHz and $ERBnumber$ the number of ERB.

Three impacts linked to the latch part are then detected. Actually, the three most energetic impacts are separated by at least 13 ms (i.e. the time needed to perceptually separate them) by calculating level vs. time functions from high frequency sub-bands. These impacts correspond to the main latch mechanism impacts and eventually the door/seal impact. The main impact related to the low frequency resonance part is also identified. The levels of the ERB sub-band signals are calculated around the time impacts and their damping laws are estimated from the analytic signals in each sub-band. The temporal envelope is assimilated to a decaying exponential function e^{-at} as often assumed for percussive sounds. The damping laws are characterized by a in each ERB sub-band associated to a weight calculated from the mean level in the sub-band. In addition, two exponentially damped sinusoids are extracted from the low frequency part (<70Hz) or the resonance part with the Steiglitz-MacBride method [11]. Those are the only modes considered in our model. For higher frequencies we consider each ERB sub-band as noise.

The levels and damping factors in each critical band are then approximated in order to reduce the number of parameters of our model. All these approximations were found by analyzing real sounds, listening to corresponding synthesized sounds and specifically sensory descriptors, adjusting the number of parameters and reiterating the process. For the impacts of the latch part, the level laws are approximated with 2nd order polynomial functions, and the damping laws are approximated with a constant for the two first clicks and 2 constants for the last click, since from a perceptual point of view this click is the most salient and therefore necessitates a more precise damping adjustment. The minimal damping law is applied to ERB bands with a strong damping factor weight. For the resonance part, the level law is approximated by 2 straight lines, and the damping law is approximated by an exponential function (Fig. 1).

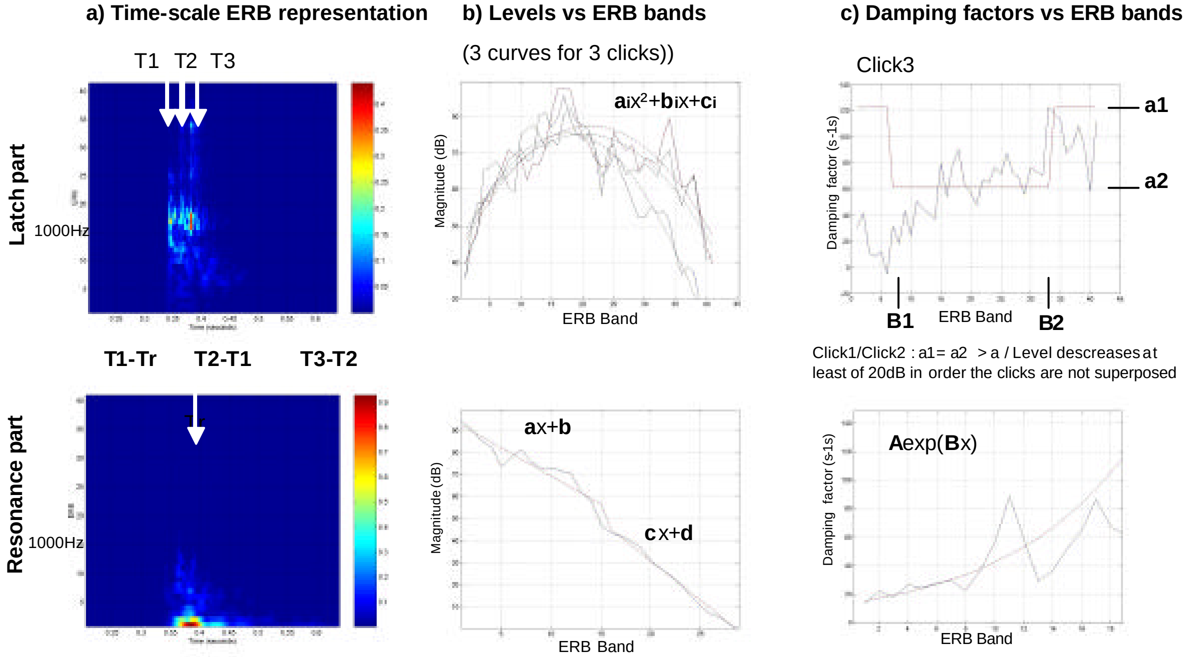

Figure 1.- Analysis - signal parameters (in bold). The upper part of the figure corresponds to the latch part of the signal, while the lower part corresponds to the resonant part of the signal. (a) Time-scale ERB decomposition (recombination of EMD modes). Detection of main impacts. (b) Levels vs ERB bands measured and approximated (dashed curves) for the 3 clicks of the latch part and for the impact of the resonance part. (c) Damping factors vs ERB bands measured and approximated (dashed curves).

Synthesis

4 white noises are filtered by the Gaussian filter bank. Each sub-band signal is adjusted to gain factors (from approximated curves), and multiplied by temporal envelopes which are exponentially ramped (fixed) and damped (damping from approximated curves). The sub-band signals are then added to form the impact. Two exponentially damped modes are added to the resonance part. Each impact starts at the right time depending on the time analysis parameters. The synthesized signal is obtained by addition of all the synthesized impacts.
We implemented the synthesized model in real-time (Fig.2), using the Max-MSP software. The model is controlled by 26 analysis parameters (22 described above and 4 modal parameters). The interface makes it possible to listen to the latch part, the resonance part, and the whole signal.

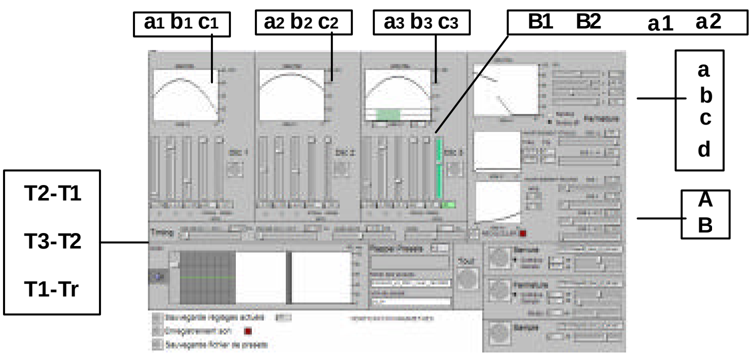

Figure 2.- Real-time synthesis interface controlled by signal analysis parameters

Thanks to this real-time tool, we can manipulate and improve the analysis-synthesis model via expert listening, with a particular attention to the relevant sensory properties. We further tempt to validate the analysis-synthesis model via perceptual experiments on resynthesized sounds.

PERCEPTUAL ANALYSIS OF RE SYNTHESIZED SOUNDS

The same 26 door closure sounds are synthesized. The validity of the synthesis is controlled by perceptual analysis. We conducted similarity ratings to observe global perceptual distances between real and synthesized sounds, and acoustical sensory profile experiments, here used as our reference.

Similarity ratings

We first wanted to observe if the synthesized sounds are perceptually close to their homologous real sounds. To this aim, we chose 6 sounds that differed by their values of main sensory descriptors. These 6 door closure sounds were very different and their homologous synthesized sounds are used for the experiment. 60 subjects without major hearing loss were recruited, where 60% were men and 40% were women, 25% < 35 years old, 50% < 50 years old, 25% > 50 years old. They were paid for their participation. Similarity ratings between all pairs of stimuli were computed.

We analyzed similarity scores via Hierarchical Ascendant Classification (HAC) and obtained 3 groups which contained 2 sub groups of real sounds (p3_s) and their homologous synthesized sounds (SYNTHES_p3_s) (Fig.3)

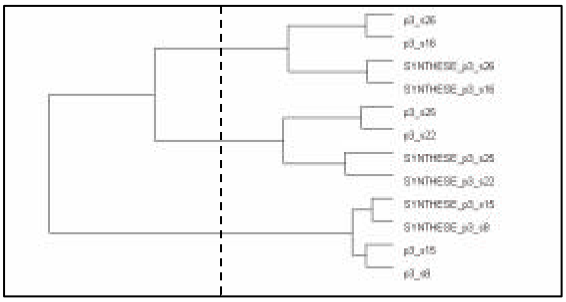

Figure 3.- Hierarchical Ascendant Classification of similarity ratings

The analysis-synthesis model makes it possible to reproduce door closure sounds within gross categories. Moreover no naive subject expressed doubt about the realism of the synthesized sounds.

Acoustical sensory profiles

10 subjects (same as in the first experiment) established the sensory profile of 26 synthesized door closure sounds. The sounds for which the descriptors have maximal and minimal values were chosen as reference. Discriminability, repeatability and consensus were verified (as *below*).

We compared the evaluations of real and synthesized sounds (Fig.4). LOUDNESS evaluations are quite similar (R2=0.77). BOMN evaluations allow for the classification of sounds in 2 groups. KE evaluations are always underestimated but the order of the sounds is relatively well respected except for 2 sounds (R2=0.72 without 2 sounds).

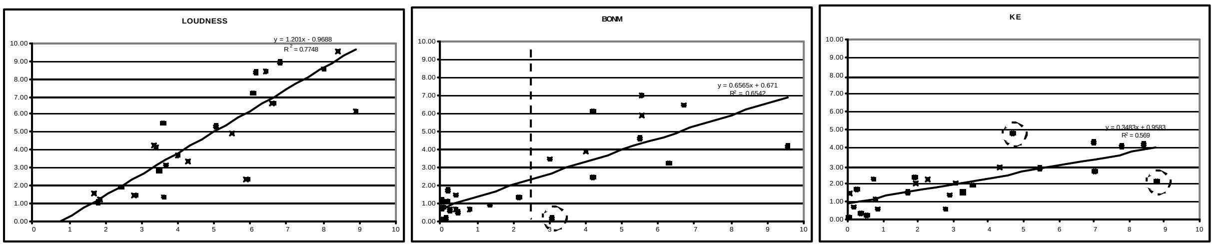

Figure 4.- Perceptual evaluations of synthesized vs. real sounds.

Sensory properties are roughly conserved by the analysis-synthesis model.

INFLUENCE OF SIGNAL PARAMETERS ON PERCEPTUAL PROPERTIES

By means of the real-time door closure synthesis tool, we listened to the effect of the 26 parameters and chose 12 parameters assumed to have a strong influence on perceptual properties. We decided to specifically study these signal parameters (P1 to P12) by creating 12

series of 9 sounds starting from the ''mean'' sound, *i.e.* the sound which all parameters corresponded to the mean values measured on real sounds. Each series is constructed by modifying only 1 parameter with a constant step from the minimal value to the maximal value measured on real sounds (M1: min value to M9: max value).
All sounds were evaluated by the 10 trained subjects on the sensory descriptors LOUDNESS, BOMN, and KE. Repeatability and consensus were verified (see below). We present some results for the descriptor KE to illustrate the interest of our analysis-synthesis approach (Fig.5). We observed for example the effect on parameter P2 (Level of one click) and examined the effects of signal parameters by comparing Mean Squared Errors (MSE).

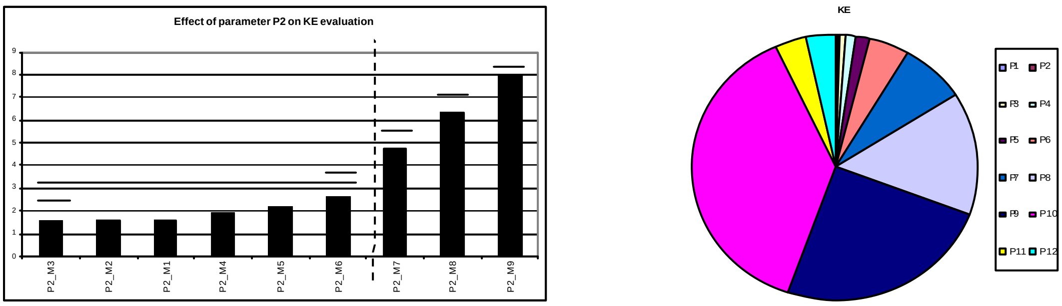

Figure 5.- (a) Effect of parameter P2. (b) Effects of signal parameters (%MSE).

This result highlights the non linear effect of P2 on KE perception and in particular reveals a threshold (M6/M7) from which this parameter has a strong effect. Moreover we observe the main signal parameters that perceptually modified the ''mean'' door closure sound. The same type of results is observed for other sensory descriptors and other parameters.

CONCLUSIONS

We here proposed an analysis-synthesis model based on perceptual studies (sensory descriptors). The usability of our model was confirmed by two experiments: a dissimilarity experiment that shows that some real and homologous synthesis sounds are perceived similarly within gross categories of door closure sounds; the acoustical profiles of real and synthesized sounds are close. The model allows for a synthesis of sounds recognized as door closure sounds that roughly conserves the main sensory properties. We conducted experiments with synthesis sounds created by tuning 12 signal parameters to observe the main effects of these parameters on the sensory descriptors (thresholds, main parameters).
As a next step, we plan to realize other experiments to study the interactions between parameters to further establish predicting models of door closure sound perception that allow for an easy interpretation. Such models should help the supplier in the evaluation of technical solution design to improve door closure sound quality.

References: [1] E. Parizet, E. Guyader, V. Nosulenko: Analysis of car door closing sound quality. Applied Acoustics, In Press.
[2] V. Roussarie, M-C. Bezat: How do we hear door closure sound? To submit.
[3] A. Caclin, S. McAdams, B. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. Journal of Acoustical Society of America **118(1)** (2005) 471-482.
[4] B. L. Giordano, K. Petrini: Hardness recognition in synthetic sounds. Conference Proceedings, Stockholm, Sweden SMAC03 (2003)
[5] S. McAdams, A. Chaigne, V. Roussarie: Psychomechanics of simulated sound sources. Journal of Acoustical Society of America **115(3)** (2004) 1306-1320.
[6] M. Arakami, R. Kronland-Martinet: Analysis-synthesis of impact sound by real-time dynamic filtering. IEEE Transactions on Audio, Speech and Language Processing **14(2)** (2006) 695-705
[7] M. Aramaki, R. Kronland-Martinet, T. Voinier, S. Ystad: A percussive sound synthesizer on physical and perceptual attributes. Computer Music Journal **30:2** (2006) 32-41
[8] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung and H.H. Liu: The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. London A, **Vol. 454** (1998) 903-995
[9] P. Flandrin, G. Rilling, P. Gonçavels: Empirical Mode Decomposition as a Filter Bank. IEEE signal processing letters **11(2)** (2004) 112-114
[10] B. J. C. Moore: An introduction to the psychologiy of hearing. 4th edition.1996.
[11] K. Steiglitz, L. E. McBride: A technique for the identification of linear systems, IEEE Trans. Automatic Control, **AC-10**, (1965) 461-464