DISSIMILARITY TEST MODELLING BY TIME-FREQUENCY REPRESENTATION APPLIED TO ENGINE SOUD

JEAN-FRANÇOIS SCIABICA[1], ANAIK OLIVERO[2], VINCENT ROUSSARIE[1], SØLVI YSTAD[2] AND RICHARD KRONLAND-MARTINET[2]

[1] *PSA Peugeot-Citroën, Paris, France*
jeanfrancois.sciabica@ext.mpsa.com, vincent.roussarie@mpsa.com
[2] *Laboratoire de mécanique et d'acoustique (LMA), Marseille, France*
{olivero, ystad, kronland}@lma.cnrs-mrs.fr

Listening tests like dissimilarity tests are traditionally used to study sound perception and build perceptive timbre spaces. As the processing of such tests is tedious when a large number of sounds are to be judged, this work proposes a method to automatically compute the dissimilarity between sounds by combining an auditory representation, the cochleagram, and the concept of time-frequency masks. The cochleagram gives an energy distribution in a time-frequency space which can be considered as a perceptive sound representation. By introducing auditory representations in time-frequency analysis, we wish to propose new tools to characterize timbre. Time-frequency masks that carry information about the time-frequency differences between two signals applied to auditory representations might lead to global timbre descriptors that, as opposed to most traditional timbre descriptors, also are adapted to signals that strongly evolve with time. In this study we applied the proposed method to engine sounds that are complex and rich signals whose perception strongly depends on the dynamic sound variation. In order to evaluate the robustness of our model, we compared the timbre space obtained with perceptive tests to the one generated by our proposed method.

INTRODUCTION

Perceptive description of interior car sounds is an important challenge for car manufacturers because such sounds influence the identity and the perceived quality of the car. Acousticians can indeed conceive and design sporty engine sounds in order to increase the sensation of acceleration. Perceptive studies are therefore carried out to evaluate such sounds and to identify perceptually relevant signal parameters. The identification of these parameters is essential to conceive specific car sounds that improve the impression of the car.

Sounds perceived in car passenger compartments are the result of three acoustic sources: the engine sound, the tire-road source and the aerodynamic source. The resulting signal is therefore a mixture of several harmonics and a low-frequency broad band noise. The engine sound is a complex sound rich in overtones. Its fundamental frequency varies with the engine rotation speed, and the level of each harmonic depends on the multiple resonances inside the car. In addition, masking phenomena [1] must be considered to identify audible harmonics in the noise.

The main dimension that characterizes interior car sounds is the audibility of engine sound, which is mainly correlated to the engine sound loudness [2]. When the engine sound is sufficiently audible in the car, it can be described by perceptual attributes such as booming, brightness and roughness. Booming is associated with a resonant low-frequency harmonic and can be considered as annoying for the driver [3]. Increased brightness reflects the presence of audible high-order harmonics, while increased roughness reflects audible secondary harmonics that interact with the main harmonics. Although these perceptual attributes can be clearly identified at a given instant, they fail to properly characterize the dynamic variation of the car sounds, for instance during acceleration. Timbre descriptors that take into account both time and frequency variations are needed for this purpose.

As a first approach we therefore propose to work with the so-called cochleagram [6] which corresponds to the output of an auditory model [5] based on physiological studies of the ear. The cochleagram displays the perceived energy distribution of a sound in a time-frequency space and enables the identification of audible signal components. Applied to an engine sound, the cochleagram shows the engine harmonics that contribute to the perceptive attributes of the sound [7]. In this study we show that working with time-frequency images makes it possible to take into account the signal features (brightness, roughness…) and their evolution over time, which is an important perceptual cue of non stationary signals such as interior car sounds.
A major question that occurs when working with global timbre attributes is how such perceptually relevant time-frequency representations further can be used to compare and categorize different engine sounds. In

traditional timbre studies, listening tests are carried out to obtain dissimilarity judgements between pairs of sounds. Dissimilarity scores are then placed in a dissimilarity matrix which is used in a Multidimensional Scaling (MDS) representation to obtain a perceptive space. Generally, 2 or 3 dimensions are enough to describe the perception and to give a visual representation.

It could therefore be interesting to extract dissimilarity ratings directly from our perceptually relevant time-frequency representations. This would give us a global time-frequency representation of the dissimilarity between two sounds and enable us to avoid listening tests that require a lot of time and a large number of sounds. Such a strategy was applied for the characterisation of loudspeakers by considering Zwicker's mean specific loudness as a perceptive descriptor [4]. To estimate the difference between two time-frequency images, we here propose to use the concept of time-frequency mask (or Gabor masks).

A Gabor mask between two signals carries information about the time-frequency differences between these two signals and can be considered as a time-frequency transfer function. Gabor masks were used in [8, 9] to compare isolated notes of musical instrument. By averaging the energy contained in the Gabor mask, the authors in [8] obtained a measure of the difference between the time-frequency representations of two sounds, and they showed that this difference was sufficient to perform a good categorization of different musical instruments playing the same note. In the present context, we propose to add perceptual considerations to this approach by replacing the time-frequency representations by cochleagrams, and calculating a mask between cochleagrams, hereafter called the auditory mask. An auditory mask will then contain time-frequency information related to the perceptive differences between two sounds. In addition, it corresponds to a global timbre description because it contains the whole perceptual information. We make the hypothesis that an auditory mask provides good results in a categorization task of interior car sounds, and can also be used to predict a perceptive space.

We first explain the construction of our prediction model, by describing the auditory representation and the dissimilarity estimation derived from the auditory masks. Then, we apply this method to synthesized engine sounds obtained with the so-called HARTIS synthesizer [10] developed at Peugeot-Citroen, and engine sounds recorded in different cars. Finally, in order to evaluate the robustness of our model, we compare the timbre space obtained with perceptive tests to the one generated by our proposed method.

1 DISSIMILARITY TEST MODELLING

The dissimilarity test model is constructed by evaluating the information contained in the auditory mask of each couple of sounds. We derive an auditory mask from the cochleagrams by mimicking the formulation of a time-frequency mask. Then, this auditory mask is used to measure the dissimilarity between sounds, by properly averaging its time-frequency information.

1.1 Auditory representation

An auditory model simulates different stages of our peripheral auditory system. The transfer function of sounds through the outer and middle ear can be modelled using a single FIR filter [11]. The output of the filter can be considered as symbolizing the sound reaching the cochlea. The cochlea can be described as a bank of 4th order linear gammatone filters [12]. The frequencies of the auditory filterbank are linearly spaced on the so called ERB (Equivalent Rectangular Bandwidth) frequency scale. The step size determines the density of the filters on the ERB scale. We chose a step size of 0.1 ERB and we considered frequencies between 50 and 1200 Hz, so that 152 auditory bins were described. The next stage was the modelling of the inner cells [13] by a half-wave rectification followed by a low pass filter with a 50Hz cut-off frequency. We added a compression model with a power law (order 0.3) to model the non-linearity of the ear [14]. The output of our model can be viewed as the cochleagram defined by Slaney et al. [6] and gives a representation of the perceived energy in the time-frequency domain.

1.2 The auditory mask

We here recall the estimation problem of a Gabor mask, from which we derive an expression for the auditory mask. Time-frequency representations are invertible signal representations. More precisely, they are images of complex numbers containing particular correlations between their coefficients, and are used in the context of analysis/synthesis of signals. A time-frequency multiplier acts on a signal by multiplying its time-frequency representation with a time-frequency transfer function, called the Gabor mask. As studied in [15], a Gabor mask can be estimated between two signals, and is given as the solution of a regularized least-square problem in the signal domain. This problem is difficult to handle and can be simplified, when we assume that the time-frequency coefficients are independent.

This also leads to an expression of the problem in the Gabor domain, where a Gabor mask is defined as follows:

$$\mathbf{m}_{Gabor}^{(0,1)} = \arg\min_m \left(\left\| X_1 - mX_0 \right\|_2^2 + \lambda \left\| |m| - 1 \right\|_2^2 \right) \quad (1)$$

where X_i is the time-frequency representation of the signal x_i, λ is a regularization parameter, and the multiplication $\mathbf{m}X_0$ is performed component-wise. The parameter λ prevents from the division by zero, and controls the deviation from 1 of the Gabor mask coefficients, as the value 1 corresponds to "no transformation". By differentiating Equation (1) with respect to the modulus and the argument of **m**, the solution is obviously obtained as

$$\mathbf{m}^{(0,1)}_{Gabor} = \frac{S_1 S_0 + \lambda}{S_0^2 + \lambda} e^{i \arg(\overline{X_0} X_1)} \qquad (2)$$

where S_i is the modulus of X_i, and multiplications and divisions are performed component-wise. The Gabor mask is a complex matrix and can be used to reconstruct x_1 from the time-frequency matrix $\mathbf{m}X_0$, for a well-chosen value of λ(small comparing to S_0^2).

The auditory mask is obtained by mimicking the Gabor mask modulus defined by Equation 2. If C_0 and C_1 are the cochleagrams corresponding respectively to the signals x_0 and x_1, we define an auditory mask between x_0 and x_1 as:

$$\mathbf{m}^{(0,1)}_{Auditory} = \frac{C_1 C_0 + \lambda}{C_0^2 + \lambda} \qquad (3)$$

In this work, we chose $\lambda = 10^{-12}$. It is worth noticing that a mask is a time-frequency matrix, and the more its coefficients deviate from 1, the more the signals are different. Therefore, in order to capture relevant time-frequency information, the two signals must be aligned in the time-frequency domain. Indeed, a time-frequency shift between C_0 and C_1 will generate large values in the Gabor mask, which are not relevant in a timbre categorization task. As we want to describe subtle differences between two signals, we consider signals with the same length and the same fundamental frequency variation over time. A processing [16] can be applied to sounds which do not respect the temporal condition. Nevertheless, we assume that our method can be only applied to real-life sounds with the same fundamental frequency.

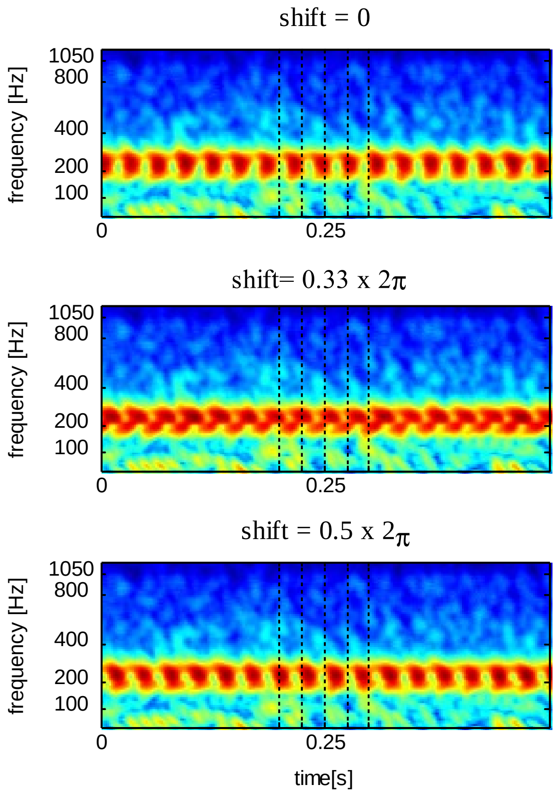

Figure 1: Impact on the cochleagram of phase shift between three beating engine harmonics. First and third harmonics are in phase while second harmonic is phase shifted. Cochleagram are show for different shift.

In addition, a phase shift between signal harmonics can be visualized on the cochleagram [17]. In the case of beating harmonics, the perceived energy modulation associated to roughness depends on this phase shift. Given a harmonic signal, the same signal with different phase shifts for its harmonics will sound as rough as the initial one (see Fig. 1). However, on the cochleagrams, the dips and the peaks will not coincide due to these phase shifts. In this situation, the auditory mask between these two signals will deviate from the unity mask while the signals sound the same. We propose to estimate a phase shift between each bin of the two cochleagrams by maximizing the correlation product. We also built two new cochleagrams whose perceived energy modulation matches. Finally, the cochleagrams are adjusted so that peaks and dips coincide.

1.3 Perceptive space modeling

As the auditory mask contains a time-frequency information of the dissimilarity between x_1 and x_0, the Euclidean norm $\|\mathbf{m}^{(0,1)}-1\|$ will give us a global measure of the dissimilarity between the two signals. However, it has been showed in [8] that the following symmetrised

Itakura-Saito divergence between the mask and 1 gives better results. This divergence is obtained as:

$$d^{(0,1)} = \frac{1}{2}[\|m^{(0,1)}\|_1 + \|m^{(1,0)}\|_1 - \|\log|m^{(0,1)}|\|_1 - \|\log|m^{(1,0)}|\|_1 - 2] \tag{4}$$

We can assume that the presence of the logarithm in the divergence makes it possible to take into account the logarithmic perception of the intensity of sounds.

In our experiments, we considered N interior car sounds, computed an auditory mask for each pair of signals, adjusted their phase shift, and derived a measure of the dissimilarity between sounds with Equation 4. Finally, we obtained a dissimilarity matrix for the N sounds. Then, a multidimensional scaling representation was obtained from the dissimilarity matrix, and compared with the perspective space obtained by listening experiments. This comparison is described hereafter.

1.4 Robustness evaluation method

We propose to evaluate the robustness of the dissimilarity model by comparing the results of our estimation with the results of a listening test. A procrustean analysis allows for comparisons between these two spaces by finding the best way to match the two spaces only with geometrical transformations. The matching error is computed with a goodness of fit indicator (also called the Procrustes distance), defined as the squared Euclidian distance between the positions of each point in the two spaces. The lower is the goodness of fit; the better is the matching between the two spaces.

2 APPLICATION TO ENGINE SOUND CARACTERIZATION

We applied this method to interior car sounds because their fundamental frequency evolves slowly over time and their harmonics level varies over time. Timbre descriptors built on stationary sounds fail to predict engine sound perception. We considered two kinds of sounds: synthesized sounds whose perceptive attributes are well controlled, and recorded sounds from 12 different accelerating cars.

2.1 Stimuli

2.1.1 Synthesized sound

12 interior car sounds were synthesized with the HARTIS engine synthesizer used at PSA Peugeot Citroën. This software allows a real time control of the perceptive parameters. 4 brightness levels and 3 roughness levels were used to create 12 stimuli with a factorial experimental design. The duration of each sound was fixed to 1.7 seconds, and each sound corresponded to an accelerating car with a rotation speed variation from 3000 to 4500 rotations per minute. These sounds were equalized in loudness during a pre-test. Then, they were evaluated during a dissimilarity test by 55 subjects. These dissimilarity ratings were finally used to obtain a 2-dimensional perceptive space of these sounds, where the perceived dimensions corresponded to the brightness, and the the roughness.

2.1.2 Recorded sounds

Interior car sounds are recorded with a dummy head in 12 cars from different manufacturers during acceleration. We only conserved the part of the acceleration corresponding to a motor rotation speed between 3500 and 4300 rotations per minute. The sounds were scaled in time and lasted for 2 seconds. In addition, they all had the same rotation speed variation over time.

These sounds were perceptually evaluated by a sensory panel [18] made of trained subjects who first proposed sensory descriptors revealed by onomatopoeia to describe sounds. In our case, 9 onomatopoeia were found to describe the different noises perceived in interior car: descriptors for engine sound (e.g. “ON” associated to the sound booming and “REU”, associated to the sound roughness), descriptors for aerodynamic sound (e.g. “FF” and “SHH”) and descriptor for their dynamic evolution and their interaction. Then they proposed a scaling for each descriptor. The panel finally gave an evaluation of 12 sounds on each descriptor and a perspective space was built from this description with a factor analysis. The 2 main dimensions were the descriptor “ON” (sound booming) and REU (sound roughness).

2.2 Pre-results on engine sound perception with cochleagram

We here describe the main results obtained by applying our auditory model on these interior car sounds [7]. First, on Figure 2, the cochleagram mainly shows audible engine harmonics, as masking phenomena are taken into account in the auditory model. Secondly, the engine roughness is clearly observed on the cochleagram, by the presence of an amplitude modulation that we call the “cochleagram beat”. This amplitude modulation allows us to identify the interaction between the harmonics, which is responsible for the roughness perceived in these sounds. These beats clearly appear on Figure 1 and their beating frequency equals the frequency of the amplitude modulation. By comparing to the listening test, the more the beat is important, the rougher the engine sound is. Indeed, the amplitude modulation is directly correlated to perceived roughness [19]. Thirdly, as the cochleagram is a perceived energy representation, we can derive a brightness evolution over time.

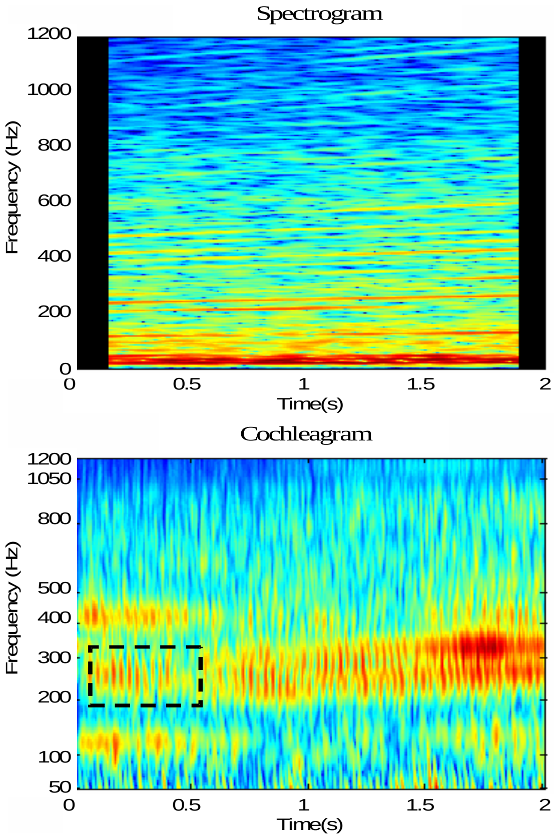

Figure 2: Spectrogram (on the top) and Cochleagram (on the bottom) of an engine sound recorded in an interior car. The beats associated with the roughness are encircled in dash-line on cochleagram.

This kind of representation is really useful in the case of engine sounds as it allows us to identify the perceptive relevance between the signals components. Furthermore this representation reduces the information quantity, as the non-audible components are not considered.

2.3 Results on the perceptive space prediction

2.3.1 Synthesized sound

A listening test is carried out to obtain dissimilarity judgements between pairs of sounds. An MDS is applied on the dissimilarity matrix to obtain a perceptive space. The two perceptual axes correspond to the synthesis parameter (brightness and roughness). Results are plotted in Figure 3. The perspective space computed with the Gabor mask is compared with the reference space in the upper part of the figure, whereas the perspective space computed with the Auditory mask is compared with the reference space I in the lower part of the figure. In each case, the two spaces are scaled with a procrustean analysis. Table 1 gives the goodness of fit issued from this analysis. It also gives the correlation product between the coordinates in each point for the two dimensions.

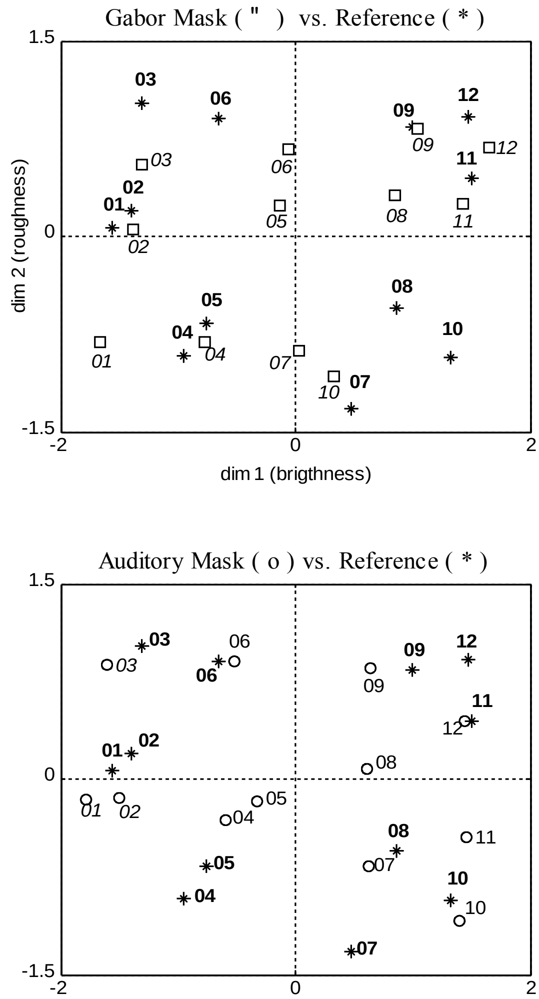

Figure 3: Comparison between the perspective space obtained from listening tests (reference space) and the perspective space computed with the Gabor mask (upper part of the figure), and the one computed with the Auditory mask (lower part of the figure).

	Gabor mask	Auditory mask
Goodness of fit	0,21	0,14
Correlation (Dim 1)	$R^2 = 0,93$	$R^2 = 0,98$
Correlation (Dim 2)	$R^2 = 0,79$	$R^2 = 0,82$

Table 1: Goodness of fit obtained from procrustean analysis and the correlation coefficients between the reference space and the models.

2.3.2 *Recorded sound*

Results are plotted in Figure 4. The two perceptual axes are obtained from a factor analysis computed on the descriptor issued from sensory analysis. The perceptive space computed with the Gabor mask is compared with the reference space in the upper part of the figure, whereas the perspective space computed with the Auditory mask is compared with the reference space in the lower part of the figure. In each case, the two spaces are scaled with a procrustean analysis.. Table 2 gives the goodness of fit issued from this analysis. It also gives the correlation product between coordinates in each point for the two dimensions.

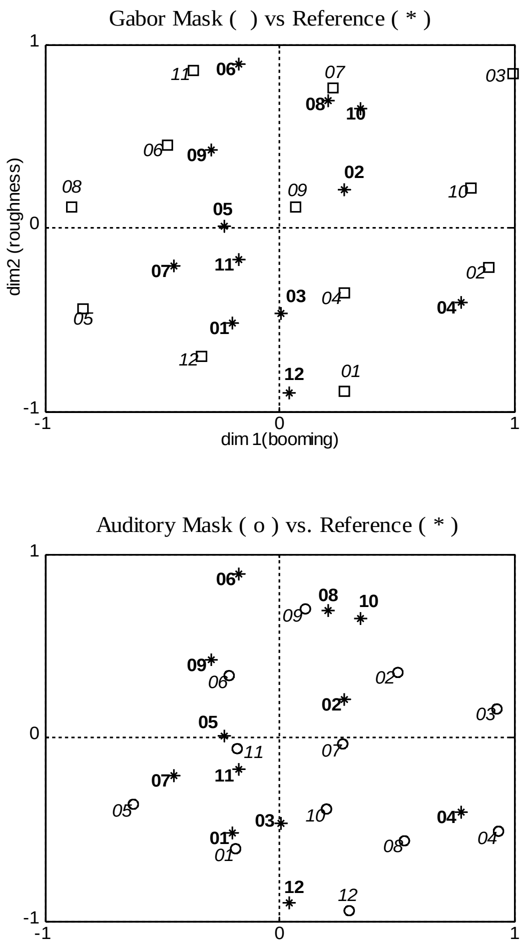

Figure 4: Comparison between the perspective space issued form listening (Reference) and the perspective space computed with Gabor mask (upper part of the figure) and with auditory mask (lower part of the figure).

	Gabor mask	Auditory mask
Goodness of fit	0,91	0,73
Correlation (Dim 1)	$R^2 = 0{,}29$	$R^2 = 0{,}64$
Correlation (Dim 2)	$R^2 = 0{,}31$	$R^2 = 0{,}46$

Table 2: Goodness of fit obtained from procrustean analysis and the correlation coefficients between the reference space and the models.

3 DISCUSSION

For the two sets of sounds, the matching between the reference space and the estimated space is better with the auditory mask than with the Gabor mask. The goodness of fit associated with the matching error is lower in both experiments for the auditory mask. Theses results confirm the interest of considering a perceived time-frequency representation to capture a realistic dissimilarity between sounds. The perceptive attributes like booming or brightness are better modelled. We also observe that, when the auditory mask is used, the correlation between space dimensions is higher for the booming (with the measured sounds) and the brightness (with the synthesized sounds). Roughness is indeed a perceptive attribute whose characterization is still debated [20], especially the different ways to sum partial roughness issued from the auditory filters (the cochleagram rows) to estimate a global roughness. Partial roughness have indeed to be weighted and summed [21]. Each cochleagram row is considered as the auditory filter output and the amplitude modulation can be considered as the partial roughness. In our method, we didn't consider weighting functions and we only compared partial roughness. This might explain the lack of precision when we compare sounds on this perceptive dimension. These weighting functions could be integrated in the mask measure defined by equation (4) in order to improve our method.

In addition, we can notice that results are better for synthesized sounds than for measured sounds. This is not surprising, as the synthesized sounds were constructed to match two particular perceptive dimensions. A common way to decide the number of perceptual dimensions that should be considered is to use the MDS Kruskal stress value. For synthesised sounds, the stress values confirm that two dimensions are sufficient to describe their perception. Oppositely, the recorded sounds are more rich and complex, and should be described by more than two perceptive dimensions. The sensory analysis used to obtain the perceptive space of the recorded sounds revealed 9 relevant descriptors. A principal component analysis shows that 3 dimension (ON, REU + an other dimensions describing brightness evolution over time) contains 80% of perceptive information. Here, we only consider the two most relevant dimensions to describe the perceptive space. A better match might have been obtained by comparing N dimensional perceptive spaces (N >2). Moreover, we computed our method by considering the entire signal. For both test, the sounds lasted for 2 seconds. But during listening tests, subjects often concentrate their attention on the last seconds of the sound to evaluate it [22]. That's why, we also computed our method by considering only the last seconds of the signal, but we observed no noticeable improvement

In our method, we point out the importance of cochleagram phases and propose a method to correct phase shift and scale cochleagram in phase. As the correlation product is computed on the whole signal length, we supposed constant phase shift over time. This approximation is still valid in the case of synthesized sounds, but it can also explain the results for recorded sounds. A phase scaling by window could be a way to correct this approximation, but it might be complex to compute by conserving time scaling.

4 CONCLUSION

In this paper, we propose a method to model a dissimilarity test by comparing the time-frequency representations of sounds. Two time-frequency representations have been tested; the time-frequency representations and the cochleagram, a perceived energy time-frequency representation issued from an auditory model. The latter enables to emphasize the perceptive attributes in the time-frequency domain. We have applied our method to interior car sounds, as their signals structure is suitable for this time-frequency comparison and because it's a complex and rich signal whose perceptive description depends on the dynamic variations of the sound. We show that auditory representation improves the computation of dissimilarity tests for different kinds of tested sounds (synthesized or recorded sounds). Progress on roughness characterization will certainly improve this dissimilarity calculus test and an application to music instruments, which perspective spaces are well known [23] will enable a better calibration of this method.

ACKNOWLEDGMENTS

This project has been partly supported by the French National Research Agency (ANR-10-CORD-010, "Métaphores sonores", http://metason.cnrs-mrs.fr/)

REFERENCES

[1] F. Richard, F. Costes, J-F. Sciabica, V. Roussarie, "Vehicle Acoustic specifications using masking models", *Proceedings of the Internoise,* Istanbul. (2007)

[2] P. Susini, S. McAdams, "Loudness asymmetry ratings between accelerating and decelerating car sound", *Proceedings of the Acoustics'08 conference*, Paris. (2008)

[3] H. Hansen, R. Weber, U. Letens, "Quantifying tonal phenomena in interior car sound", *Proceedings of the forum acusticum*, Budapest. (2005)

[4] M. Lavandier, S. Meunier, P. Herzog, "Identification of some perceptual dimensions underlying loudspeaker dissimilarities", *J. Acoust. Soc. Am*., Vol 123(6), pp 4186-4198. (2008)

[5] R.D. Patterson, M.H. Allerhand, C. Giguere, "Time-domain modelling of peripheral auditory processing : A modular architecture and a software platform", *J. Acoust. Soc. Am.*,Vol 98, pp 1890-1894. (1995)

[6] M. Slaney, D. Naar, R.F. Lyon, "Auditory Model Inversion for Sound Separation", *Proceedings of the IEEE-ICASSP*, pp 77-80, Adelaide. (1994)

[7] Sciabica J.F, Bezat M.C., Roussarie V., Kronland-Martinet R., Ystad S. "Timbre Characteristics of Interior Car Sound" *Auditory Display, Springer Verlag Berlin Heidelberg*. (2010).

[8] A. Olivero, L. Daudet, R. Kronland-Martinet, B. Torrésani, "Analyse et catégorisation de sons par multiplicateurs temps-fréquence", *Actes du colloque GRETSI sur le traitement du signal et des images*, Dijon, France. (2009)

[9] Ph. Guillemain, Ch. Vergez, D. Ferrand and A. Farcy, "An instrumented saxophone mouthpiece and its use to understand how an experienced musician play", *Acta. Acustica united with Acustica* 96, pp 622-634. (2010),

[10] F. Richard, V. Roussarie, "Sound design in car passenger compartment : Process and tool for the control of engine sound character", *Actes des Journées du Design Sonore,* Paris. (2004)

[11] B. Glasberg, B.C.J. Moore, "Model of loudness applicable to time-varying sounds", *Journal of the Audio Engineering Society*, Vol 50(5), pp 331-342. (2002)

[12] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank", *Acta Acustica united with Acustica*, Vol 88, pp 433–442. (2002)

[13] R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor", *J. Acoust. Soc. Am*., Vol 79(3), pp 702–711. (1986)

[14] S. Stevens, "Concerning the Form of the Loudness Function", *J. Acoust. Soc. Am*., Vol 29, Issue 5, pp 603-606. (1957)

[15] Anaïk Olivero, Bruno Torrésani, Richard Kronland-Martinet, "A new method for gabor multi-

pliers estimation : Applications to sound morphing”, *Proceedings of the EUSIPCO*, Danemark, (2010)

[16] J. P. Bello, C. Duxbury, M. Davies et M. Sandler, IEEE Signal Processing Letters, vol. 11, no. 6, pp. 553-556. (2004)

[17] B.C.J. Moore, “Interference effects and phase sensitivity in hearing “, *Phil. Trans. R. Soc. A* 360, pp 833-858. (2002)

[18] V. Roussarie, F. Richard, M-C. Bezat, “Perceptive qualification of engine sound character, validation of auditory attributes using analysis-synthesis method “, *Proceedings of the CFA/DAGA*, Strasbourg. (2004)

[19] P. Daniel, R. Weber, "Psychoacoustical Roughness: Implementation of an Optimized Model", *Acustica* 83, pp 113-123. (1997)

[20] R. Mores, T. Smit and J.-M. Wiese, "Perceived Roughness - a Recent Psycho-Acoustic Measurement", Proceedings of the 126th AES convention, Munich, 2009

[21] W. Aures, “Ein Berechnungsverfahren der Rauhigkeit“, *Acustica*, Vol 58, pp 268-280. (1985)

[22] S. McAdams, P. Susini, N. Misdariis, S. Winsberg, “Multidimensional characterisation of perceptual and preference judgements of vehicle and environmental noises “,*Proceedings of Euro-Noise 98*, Mu nich, (1998)

[23] J. Grey, "Multidimensional Perceptual Scaling of Musical Timbres". *J. Acoust. Soc. Am*., 61(5):1270–77. (1977)