

COVER PAGE

Document downloaded by @DAEL

Fri Nov 27 10:31:34 2026
For personal use

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

Bibliographical reference

Exploring Roughness Perception in Car Engine Noises through Complex Cepstrum Analysis, A. Camacho, G. Piñero, M. De Diego and A. González, Acta Acustica **vol. 94** (Number 1), 2008, pp. 130-140

DOI
https://doi.org/10.3813/AAA.918015

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 94 (2008) 130 – 140
DOI 10.3813/AAA.918015

Exploring Roughness Perception in Car Engine Noises through Complex Cepstrum Analysis

A. Camacho, G. Piñero, M. de Diego, A. González
Institute of Telecommunications and Multimedia Applications (iTEAM), Universidad Politécnica de Valencia, Spain. {acamacho,gpinyero}@dcom.upv.es

Summary
Car engine noises have been analyzed from a psycho-acoustical point of view since early 90's. Classical psychoacoustic parameters as loudness, sharpness and roughness have been used for this task, resulting in significant improvements in the annoyance caused in humans, especially for diesel engine noises. However, roughness sensation has been well defined and measured in modulated narrowband signals, but not in broadband noises. Here a new model of engine noise production is proposed in order to relate some of its features with the roughness perceived. The model is based on complex cepstrum analysis of the real engine noise, a technique widely used in speech processing, but not explored yet for psycho-acoustic sensations. A noise synthesizer is build for performing subjective tests and detect model parameters related to roughness perception. Results have shown that peak amplitudes range (represented by a certain amplitude modulation index of the impulses) is linearly related to roughness perceived, whereas frequency modulation index of the same impulse train has a weak influence on roughness sensation.

PACS no. 43.60.-c, 43.66.Lj, 43.60.Uv, 43.58.Jq

1. Introduction

Noise produced by electronic devices and cars is becoming a relevant factor for new products design process. Not only must be engineer efforts addressed to reduce sound pressure levels, but also to decrease the annoyance sensation perceived by the customer. Moreover, some product noises are designed to achieve a certain feature, that is, are made in order to sound the way they are expected to sound. Think for example in a phone bell or in a sporty car engine.

However, when evaluating the sensations produced by noises, the standard Sound Pressure Level (SPL) and other physical parameters based on simple spectral analysis are inadequate to explain the sensations perceived. By this reason, parameters related with human hearing perception, known as psychoacoustic parameters, have been widely used for this purpose.

Psychoacoustic parameters can be calculated by a formula and try to grade numerically some psychoacoustic features of the perceived sound. Some of the most important psychoacoustic parameters are:

- Loudness: it measures the sensation of sound level or power. Its calculation was standardized by the International Standard Organization in 1975 [1].
- Sharpness: it evaluates the subjective sensation of the sound timbre, mainly based on the frequency content.
- Roughness: it quantifies the unpleasantness produced by fast temporal variations of sounds.
- Fluctuation Strength: it also measures the sensation produced by temporal variations, but only by those so slow that can be followed by the human ear.

In [2], Zwicker and Fastl have described all these sensations and have carried out a lot of comprehensive experiments to analyze them, including the proposal of useful mathematical models for their calculation. These mathematical models were summarized by Fastl in [3]. They are widely used in sound quality experiments and have been programmed in several commercial software packages.

Nevertheless the psychoacoustic parameters proposed by Zwicker and Fastl were derived from experiments carried out with synthetic sounds and can provide inaccurate results when real sounds are involved. By this reason, new methods considering the ear physiology in depth were proposed by other authors, as Aures [4], Moore [5] and Blauert [6].

Regarding psychoacoustic parameters that describe sensations caused by sound envelope variations, they have been defined in the literature with the aid of an amplitude-modulated (AM) tone. If the modulation frequency of a 1 kHz tone 100% modulated in amplitude is varied from small values to higher ones, three different sensations are perceived. For very low modulation frequencies loudness goes up and down slowly, and a sensation of sound fluctuation can be perceived. The parameter used to measure it is the fluctuation strength. Going further through the modulation frequency range (over 15 Hz values) a new sensa-

Received 7 April 2006, revised 31 May 2007,
accepted 30 July 2007.

© S. Hirzel Verlag · EAA

tion is produced called roughness, whose maximum value is achieved for modulation frequencies of 70 Hz. Finally, if the frequency of modulation rises over 300 Hz, the sensation of roughness disappears and three different tones sounding independently can be distinguished.

Focusing on roughness sensation, the earliest studies about the matter came from music theory when Helmholtz [7] defined the term dissonance (unpleasant phenomenon produced when two sounds with similar frequencies are added) in contrast with consonance. Roughness was later analyzed by Terhardt [8] and several methods for its calculation were proposed by Aures [4], Zwicker and Fastl [2], and Daniel [9]. All of these methods are based on the frequency decomposition that takes place in the human ear: Sounds are split in several bands (usually the critical ones) and different calculations are made to estimate the roughness produced by each of the bands, obtaining what is called specific roughness. Finally, specific contributions are summed up to obtain the total roughness value.

Another relevant feature of human ear perception considered in all roughness calculation methods is the temporal masking effect due to sound pressure level variations of modulated tones. This effect has been studied in amplitude-modulated (AM) tones, frequency-modulated (FM) tones and narrow-band modulated noises, obtaining temporal masking patterns for different modulation indexes and frequencies. The temporal variation of a temporal masking pattern can be described by the magnitude ΔL, which represents the level difference between the maxima and the minima in the temporal masking pattern. Generally speaking, this so-called temporal masking depth, ΔL, is smaller than the modulation depth of the masker's envelope due to post-masking.

Considering the roughness calculation method proposed by Zwicker et al. [2], it states that specific roughness at each critical band is proportional to the product of the temporal masking depth ΔL and the modulation frequency f_{mod}, whereas total roughness is obtained adding all the specific roughness values. Aures [4] developed another method where roughness values were calculated in a more complicated way than Zwicker in [2], but keeping the main purpose of calculating total roughness from previous specific roughness evaluation. Both methods have proved to be suitable for AM and FM modulated tones, but usually have failed to characterize roughness produced by broadband noise. In order to improve roughness description, Daniel and Weber [9] proposed a new method that weighted specific roughness according to cross correlation coefficients between adjacent bands. Further work [10] has shown that this procedure seems to characterize roughness sensation noticeably better than Zwicker in [2], not only for white and bandpass noises, but also when AM and FM tones are used.

Summarizing the performance of the main roughness calculation algorithms, they have been tested with very simple synthetic sounds as amplitude and frequency modulated tones, obtaining fairly good results. Some of the procedures have also been applied to more complex synthetic sounds, as modulated bandpass noises, but, generally speaking, they have not been used with broadband signals that do not fit AM or FM patterns.

Regarding the particular case of broadband noise generated by a car engine, some previous experiments carried out on this type of noise have shown the extremely poor correlation between the calculated roughness and the perceived sensation of roughness. See for example references [11, 12, 13]. Moreover, some difficulties arise in the calculation method because broadband noises do not match AM or FM patterns, so estimating their temporal masking depth, ΔL, and modulation frequency, f_{mod}, for each critical band can be made only as an average.

In order to find a solution to the evaluation of roughness in broadband noises, a new approach to calculate and describe the perceived roughness of car engine noises is proposed in this paper. This approach is based on the fact that these noises fit extremely well the model of a fixed pattern that is repeated periodically except for some small variations in amplitude. Moreover, we will show in this paper that the basic pattern and the repeating period and amplitude can be extracted separately by means of cepstrum decomposition [14, 15], a method widely used in speech processing to estimate the pitch of an utterance. The experiments carried out throughout this work have shown that some small variations in amplitude together with the fixed pattern waveform are the main features that essentially affect the roughness sensation produced by this kind of noises.

We have validated the use of cepstrum analysis and synthesis by means of an experiment where synthesized noises with different levels of roughness have been generated. The level of roughness has been subjectively evaluated by means of jury tests in order to relate some relevant parameters of the cepstrum model to the roughness sensation. Results have shown that the so-called modulation index and modulation frequency were linearly related to the perceived roughness.

The outline of the paper is as follows: section 2 introduces a rather simplified model of car engine noise production, whereas section 3 explains the complex cepstrum analysis of this type of noises. Once the cepstrum analysis is explained, section 4 makes use of the parameters obtained to propose a model for synthesizing car engine noises, with the advantage that roughness level can be easily controlled through two elements: the modulation index and the modulation frequency. section 5 presents the experiments carried out to validate the synthesizer model and examines the results obtained, whereas section 6 summarizes the main conclusions of this work.

2. Model Of Engine Noise

The noise produced by a car engine working at N_{rpm} r.p.m. is nearly but not strictly periodic due to the fact that the engine cylinder suffers from small random differences in pressure between successive cycles [16, 17]. Therefore,

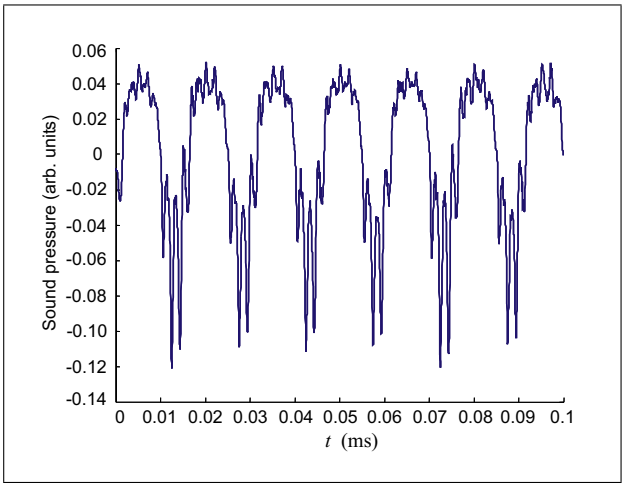


Figure 1. Noise waveform generated by a four-cylinder engine working at 2000 r.p.m.

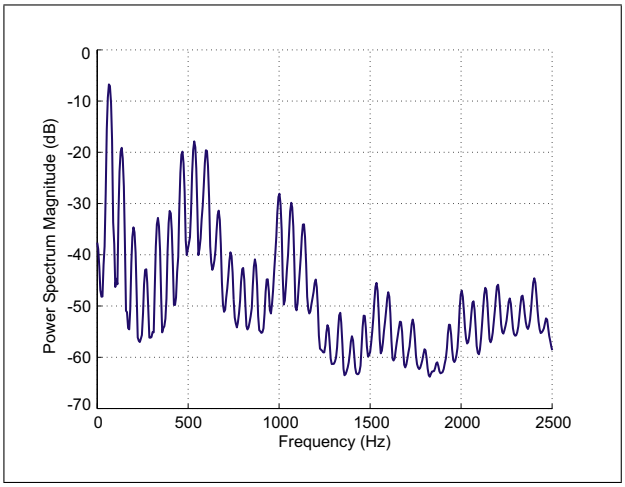


Figure 2. Power Spectrum Magnitude of the noise waveform of Figure 1.

the pressure inside the cylinders has a quasi-periodic pattern whose fundamental frequency f_0 is half the revolution speed (one explosion occurs every two engine revolutions),

$$f_0 = \frac{N_{\text{rpm}}}{120}\,\text{Hz}. \tag{1}$$

The pressure signal depends on cylinder design (valve timing, valve overlap, compression ratio, etc) and operational parameters (spark timing, engine coolant temperature, engine speed, engine load, air-fuel ratio, exhaust gas recirculation, etc) [16, 17, 18]. The movement of the engine is produced by explosions inside its cylinders that take place when it is full of inflammable gas and a spark makes it burn. Therefore, the perceived noise is a combination of combustion noise generated inside the cylinders and mechanical noise due to the engine vibrations. The quasi-periodic feature of this kind of noise can be appreciated in Figures 1 and 2, where Figure 1 shows the noise waveform produced by a four-cylinder engine working at 2000 r.p.m., and Figure 2 is a plot of its Power Spectral Density [19].

When the engine is composed of more than one cylinder, the former frequency (1) is linearly increased with respect to the number of cylinders, N_{cyl}, as follows:

$$f_0 = \frac{N_{\text{rpm}}}{120} \times N_{\text{cyl}}\,\text{Hz}. \tag{2}$$

Regarding the process of noise production explained above, Figure 3 sketches a simplified model that generates an isolated explosion inside a cylinder at time T_0. In this Figure, $x(t)$ corresponds to the pressure signal inside a cylinder when an explosion occurs in time $t = 0$, and $x(t - T_0)$ represents the same pressure signal when explosion occurs in $t = T_0$. Explosion signal has been modelled as an impulse (Dirac delta) that excites a linear and time-invariant system whose impulse response is $x(t)$, which

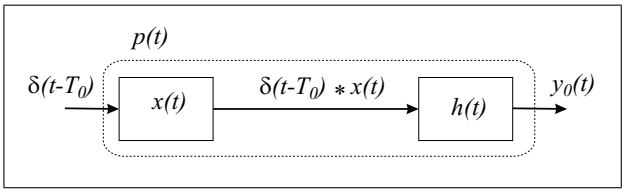


Figure 3. Model of the sound produced by an isolated explosion inside the cylinder.

according to systems theory, can be calculated as the convolution

$$x(t - T_0) = \delta(t - T_0) * x(t). \tag{3}$$

Therefore, in this model $x(t)$ has the information of the shape of the pressure wave inside the cylinder. On the other hand, $h(t)$ corresponds to the impulse response of the transmission media, which models the behavior of the mechanical system between the point where the explosion occurs (cylinder) and the point where the sound is heard. This sound can be calculated as:

$$y_0(t) = \big(\delta(t - T_0) * x(t)\big) * h(t) \tag{4}$$

and using the associative property of convolution,

$$\begin{aligned} y_0(t) &= \big(\delta(t - T_0) * x(t)\big) * h(t) \\ &= \delta(t - T_0) * \big(x(t) * h(t)\big) = \delta(t - T_0) * p(t) \end{aligned} \tag{5}$$

being

$$p(t) = x(t) * h(t). \tag{6}$$

According to linear systems theory, both systems appearing in Figure 3 could be replaced by just one with impulse response $p(t)$, that represents the sound pressure level measured at the hearing point when an explosion is produced inside the cylinder.

The sound produced by an ideal engine working at constant revolution speed, where all the explosions happened

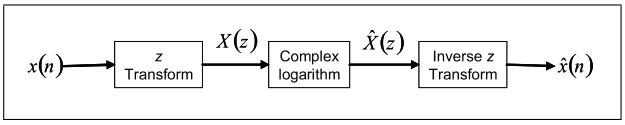


Figure 4. Block diagram of cepstrum decomposition of signal $x(t)$.

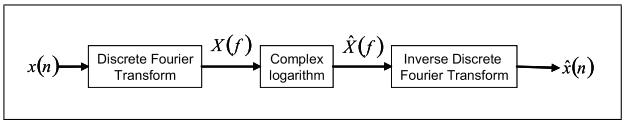


Figure 5. Block diagram of cepstrum decomposition of sampled signal $x(n)$ through the Discrete Fourier Transform.

at regular time intervals $T = 1/f_0$, would correspond to the model showed in Figure 3, but substituting the isolated impulse by a train of periodic impulses,

$$c(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT). \tag{7}$$

However, the excitation model for a real engine must consider small random variations affecting amplitudes and delays between successive explosions, so expression of the input signal $c(t)$ given in (7) should be substituted by

$$c(t) = \sum_{k=-\infty}^{\infty} a_k \delta(t - kT_k), \tag{8}$$

being a_k the amplitude of the k-th cycle, and T_k the time run between start of cycle k and start of cycle $k-1$. Therefore the mean value of T_k can be considered as the nominal period of the signal, T.

Assuming that noise generation is due to a linear system $p(t)$ that is excited by the quasi-periodic impulse train $c(t)$ of (8), the engine noise perceived by a person can be expressed as

$$y(t) = p(t) * c(t). \tag{9}$$

Our hypothesis is that roughness sensation is related to some parameters of this simplified model. Our proposal is to extract those parameters from real signals by means of the complex cepstrum analysis, and then combine them in order to synthesize signals with a certain amount of roughness, that is, whose roughness is controlled changing the model variables.

3. Use Of Complex Cepstrum Analysis In Engine Noises

3.1. Complex Cepstrum Decomposition

The cepstrum (or homomorphic) decomposition [14] has been widely used in digital speech processing applications, for example to extract the pitch of a voiced sound or for calculating spectral features [19].

The complex cepstrum of a sampled signal can be calculated through the block diagram shown in Figure 4: First of all the z-transform of sampled signal $x(n)$ is calculated to obtain $X(z)$. Then the complex logarithm of $X(z)$ is taken to get $\hat{X}(z)$, and finally the inverse z-transform of $\hat{X}(z)$ is recovered. The result $\hat{x}[n]$ is called the complex cepstrum of $x(n)$ even when it is usually a real sequence. This name is used to make a distinction between complex and real cepstrum, the second consisting basically in the same operations as the first one but using only the real part of $\hat{X}(z)$ to calculate $\hat{x}(n)$.

An important property of the complex cepstrum can be derived from its definition: the complex cepstrum of the convolution of two signals is the addition of their complex cepstra. To show this property let assume that

$$x(n) = x_1(n) * x_2(n), \tag{10}$$

then

$$X(z) = X_1(z) \cdot X_2(z), \tag{11}$$

and

$$\begin{aligned} \hat{X}(z) &= \log\left[X_1(z) \cdot X_2(z)\right] \\ &= \log\left[X_1(z)\right] + \log\left[X_2(z)\right] = \hat{X}_1(z) + \hat{X}_2(z). \end{aligned} \tag{12}$$

Once the inverse z-transform of (12) is calculated, we obtain

$$\hat{x}(n) = \hat{x}_1(n) + \hat{x}_2(n), \tag{13}$$

so that the complex cepstrum of the convolution of two signals results in the sum of their respective complex cepstra.

Although the complex cepstrum decomposition of signal $x(n)$ was originally proposed using the z-transform, practical implementations make use of the Discrete or Fast Fourier Transform (DFT) [19]. The new block diagram using this technique is shown in Figure 5.

3.2. Complex Cepstrum Of A Quasi-Periodic Signal

In section 2 we have modelled the engine noise using the continuous-time expression given by (9). Once the signal is converted to digital (sampled), a discrete-time model is stated

$$y(n) = p(n) * c(n). \tag{14}$$

As shown in equations (10)-(13), the complex cepstrum decomposition applied to $y(n)$ results in

$$\hat{y}(n) = \hat{p}(n) + \hat{c}(n). \tag{15}$$

Regarding signal $c(n)$, which is the sampled version of $c(t)$ in (8), if the variations of T_k around the average value of T_k, called T, are negligible and equal to an exact number of samples we can consider that (8) becomes

$$c(n) = \sum_{k=-\infty}^{\infty} a_k \delta[n - kN], \tag{16}$$

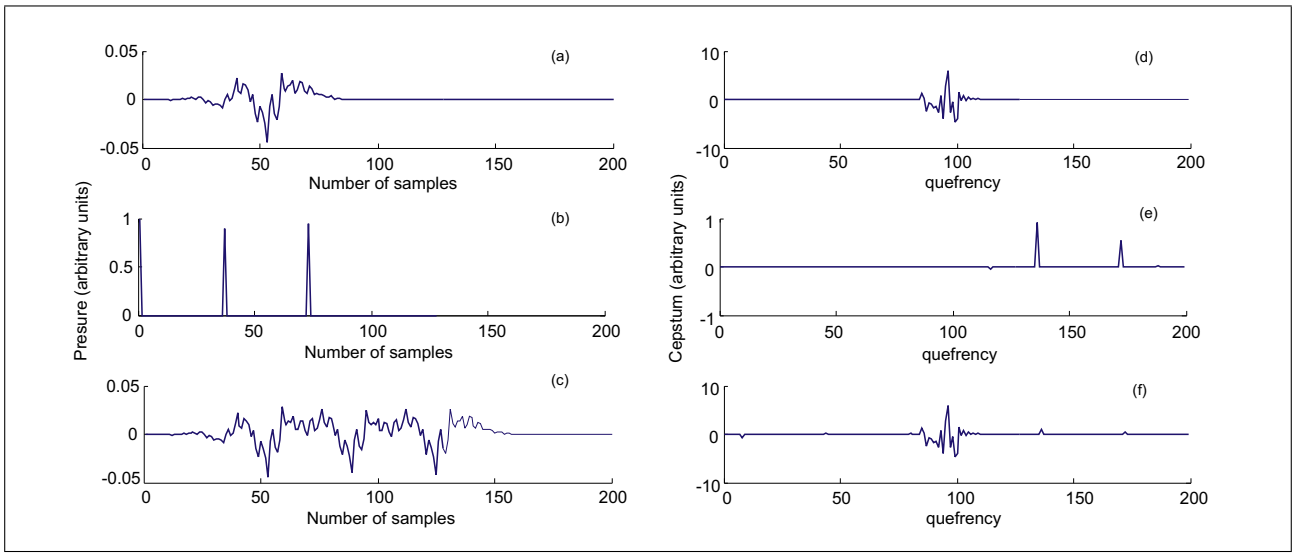


Figure 6. Decomposition of the noise signal as the convolution between a pattern $p(n)$ and three impulses $c(n)$. From top to bottom at the left: (a) pattern $p(n)$, (b) set of impulses $c(n)$ and (c) the entire noise signal $y(n) = c(n) * p(n)$. On the right side the plots (d),(e) and (f) show the complex cepstra of the signals in (a), (b) and (c) respectively.

where the only difference respect to a periodic impulse train is due to the impulse amplitudes a_k. Note that N is the period (in number of samples) of the noise waveform.

On the other hand it is not possible to work with the infinite signal $c(n)$. Actually in practical cepstrum decomposition, signal $y(n)$ is divided in intervals of a few repetitions (three or four) to avoid aliasing in the log-frequency domain [15, 19], so cepstrum processing is applied every time interval. Therefore in equation (16) only these three terms will be considered:

$$c(n) = a_0\delta(n) + a_1\delta[n-N] + a_2\delta[n-2N], \quad (17)$$

whose z-transform is

$$\begin{aligned} C(z) &= a_0 + a_1 z^{-N} + a_2 z^{-2N} \\ &= a_0\left(1-\alpha z^{-N}\right)\left(1-\beta z^{-N}\right), \end{aligned} \quad (18)$$

where α and β can be real numbers or a complex conjugate pair. Computing the logarithm of $C(z)$ we have:

$$\begin{aligned} \hat{C}(z) &= \log\left(a_0\right) + \log\left(1-\alpha z^{-N}\right) \\ &\quad + \log\left(1-\beta z^{-N}\right), \end{aligned} \quad (19)$$

and expanding the Taylor series of the logarithms we obtain

$$\begin{aligned} \hat{C}(z) &= \log(a_0) - \sum_{k=1}^{\infty}\frac{\alpha^k z^{-kN}}{n} - \sum_{k=1}^{\infty}\frac{\beta^k z^{-kN}}{n} \\ &= \log(a_0) - \sum_{k=1}^{\infty}\frac{(\alpha^k+\beta^k)z^{-kN}}{n}. \end{aligned} \quad (20)$$

Then the complex cepstrum of the impulse train $c(n)$ is

$$\hat{c}(n) = \begin{cases} \log(a_0), & n = 0, \\ -\sum_{k=1}^{\infty}\dfrac{(\alpha^k+\beta^k)}{n}\delta[n-kN], & n > 0, \end{cases} \quad (21)$$

that is a set of pulses located in time instants multiples of the signal period. As an example, Figure 6 shows three cycles of the noise produced by an engine that can be decomposed as the convolution between a pattern $p(n)$ and three impulses located at samples $n = 0$, $n = 36$ and $n = 72$ with respective amplitudes 1, 0.9 and 0.95.

As can be seen in Figure 6, the complex cepstrum of the convolution between the pattern $p(n)$ and $c(n)$ corresponds to the addition of their complex cepstra. We can state from graphs (d) and (f) that complex cepstrum of $p(n)$, $\hat{p}(n)$, can be extracted from the complex cepstrum of the total signal $y(n)$ (see equations (14)–(15)) just removing the small peaks of the right side. Note that since complex cepstrum of $c(n)$ decays at least as fast as $1/|n|$, see equation (21), other suitable method would consists in the windowing of the early samples of $\hat{y}(n)$.

3.3. Complex Cepstrum Analysis of Car Engine Noise

Summarizing the method proposed, the complex cepstrum analysis can be applied to engine noises in order to extract the pattern, $p(n)$, that will be a fundamental parameter to produce synthesized noises with controlled roughness, as it will be explained in section 4.

The procedure is as follows: First, the complex cepstrum of the real sampled noise $y(n)$ is calculated applying block diagram of Figure 5. Then, the complex cepstrum is windowed in order to isolate the information related to the pattern $p(n)$. Finally, the pattern is recovered calculating the inverse complex cepstrum.

Due to the time-varying character of engine sounds, this procedure must be applied considering only a few periods of the signal (in this work we have chosen three periods). Therefore, a new pattern is extracted every 3 periods, and the final pattern $p(n)$ is calculated as the average of all, previously aligned in time.

4. Synthesis of Engine Noises with Controlled Roughness

There exist in the literature several acoustic models for car engine noises [18, 20, 21]. Regarding the model described in [21], it was developed specifically for investigations on roughness perception. This noise synthesizer extracted the spectral characteristics of a real noise and used them in the synthesis of new engine sounds. It was based on the assumption that engine sound was generated by the sum of an amplitude modulated signal (AM) and a colored noise. Even though the model produced sounds with realistic appearance, it completely failed to describe the roughness perceived by the judges in a jury test [21].

The synthesizer of engine noise proposed in our work is based on the cepstrum analysis explained in section 3 and it can be easily described by the block diagram in Figure 7. The actual noise signal $y(n)$ is analyzed and its pattern $p(n)$ is extracted. Then, the pattern is used to filter the set of impulses $c(n)$.

Regarding signal $c(n)$ given by equation (16), peak amplitudes a_k can be chosen properly to generate amplitude modulated signals as it will be explained in section 5.2. These are the unique parameters of the synthesizer needed to control the roughness as illustrates Figure 7. On the other hand, time period N is imposed by the fundamental period T of the real engine noise, that is, the inverse of the fundamental frequency given by equation (2).

Returning to model depicted in Figure 7, a colored noise must be added to produce realistic sounds [21]. In order to generate this colored noise, the real engine signal $y(n)$ is filtered with very narrow filters to eliminate dominant harmonics. The remaining signal is then used to calculate 20 coefficients of an autoregressive system applying the autocorrelation method [22]. These coefficients are used in the all-pole filter in order to produce the required colored noise by means of white noise filtering.

5. Experiments

In order to validate the capacity of the proposed model illustrated in Figure 7 to generate synthetic sounds with controlled roughness, four jury tests have been developed to evaluate the subjective sensation of roughness for 1000, 2000, 3000 and 4000 r.p.m. synthetic sounds.

The real engine sounds $y(n)$ used as input signals in the synthesizer of Figure 7 have been digitally recorded in a semi-anechoic chamber, using a Head and Torso Simulator (HATS) placed 3 meters away from the exhaust pipe of the engine. All records have a duration of 10 seconds and have been sampled at 48000 samples/s.

The engine involved in this experiment is a commercial High-Speed Diesel Injection (HSDI) engine of a European automotive brand, and is quite representative of a sedan car noise. As it has been pointed out, the HSDI engine sounds have been recorded in a semi-anechoic chamber, using a HATS placed outside the vehicle, near the exhaust pipe end. In this sense, the recorded engine noises can be considered representative of what a pedestrian would perceive when standing near of a sedan car. Nevertheless, the general model proposed in Figure 7 can be applied to different diesel engines and different recording positions, just changing appropriately the transfer functions of the filters involved.

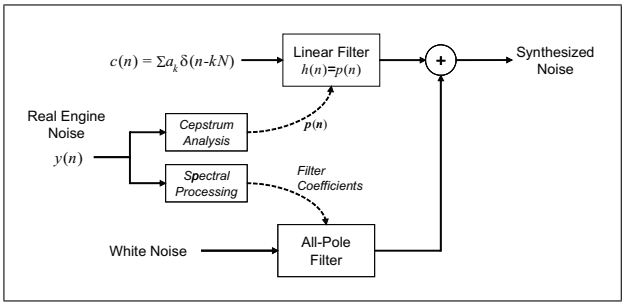


Figure 7. Proposed method to synthesize car engine noises.

5.1. Listeners

Eight subjects (7 males and 1 female) with ages from 20 to 36 years old have participated in the experiments. All of them were selected between teachers and students of the Information and Communications Technologies degree course of the Polytechnic University of Valencia. None of them reported any hearing problems when asked, neither were they specialized in acoustics or in diesel engines, so, from a certain point of view, they might be considered as potential users or costumers of this type of cars.

5.2. Stimuli

Four jury tests were designed, one for each of the engine velocities tested: 1000, 2000, 3000 and 4000 r.p.m. Each test was composed by 5 different synthesized signals generated by block diagram of Figure 7.

Regarding the generation of synthetic signals, the natural recorded engine sound was firstly analyzed using the complex cepstrum in order to extract the natural pattern $p(n)$ of (14). Then, $p(n)$ was combined with a synthetic impulse train $c(n)$ to form the sounds used in the tests. Note that in the natural engine noise signal, $c(n)$ should follow equation (16) being a_k random values, usually showing little variance. However, the synthetic impulse train is originated by the same equation, (16), but being a_k calculated by means of the following modulation scheme:

$$\begin{aligned} a_k &= 1 + m \cdot \cos(2\pi f_m t)\Big|_{t=\frac{k}{f_0}} \\ &= 1 + m \cdot \cos\left(2\pi \frac{f_m}{f_0} k\right), \end{aligned} \tag{22}$$

where m is defined as the modulation index, $0 < m \leq 1$, f_0 is the fundamental frequency of the engine sound in Hz defined in (2), and f_m stands for the modulation frequency, also in Hz.

As an example, values of m and f_m for the five signals involved in the 1000 r.p.m. test are given in Table I. Notice that modulation index and frequency values may be equal

Table I. Modulation index and frequencies used in equation (22) to generate amplitudes a_k for the five signals involved in 1000 r.p.m. test, in order to obtain synthesized noises with different roughness sensation.

Signals	Modulation Freq. (Hz)	Modulation index (%)
1	0	0
2	50	20
3	50	40
4	70	20
5	70	40

Table II. Order of comparison pairs used in the 1000 r.p.m. test. Pairs 7 and 8 are marked because they are reverse comparisons of 1 and 2.

Comparison	First signal	Second signal
1	1	2
2	3	4
3	4	5
4	3	5
5	2	5
6	2	3
7*	2	1
8*	4	3
9	1	4
10	1	5
11	1	3
12	2	4

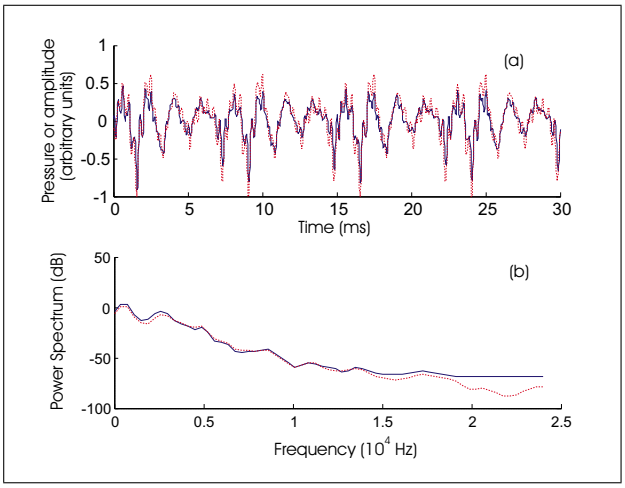


Figure 8. Comparison between a real engine noise signal (continuous) and a synthetic one (dashed).(a) Temporal waveform. (b) Power Spectrum Magnitude.

for all the tests, but that fundamental period N is related to the number of r.p.m. of the engine tested, being different for each set of equal r.p.m. noises. Therefore, synthesized engine noises with different nominal roughness have been generated for each velocity.

Figure 8 shows a real engine noise signal and its respective synthesized noise for the lowest roughness case (signal number 1 of the test). It can be observed that both signals are quite similar to each other, only showing slight differences in time between amplitude peaks and valleys.

According to theory [2], the roughness of an AM narrowband signal increases with the modulation index, reaching its maximum value at a modulation frequency of 70 Hz. However, for a broadband signal the dependence on the modulation index and frequency is not so clear. The cepstrum analysis presented in this work allows us to control the roughness with a unique parameter, a_k, depending on the modulation index and frequency by equation (22). In this sense, the main goal of the tests carried out in our experiment has been the validation of this method to describe the roughness perceived by non expert people.

5.3. Loudness equalization

It is well known that roughness is strongly dependent on the loudness level of the sounds, so in order to eliminate as much as possible this influence on the jury test, all the synthesized noises have been equalized in amplitude to present the same value of Loudness: 8 sones (equivalent to a 1 kHz tone with a Sound Pressure Level of 70 dB).

5.4. Apparatus

A Centrino Laptop PC Pentium computer running a listening test software developed by the authors [23] presented the signals to the judges and registered their assessments. The sounds were stored in the Laptop and played via a 24 bits sound card (M-Audio FireWire 410). Sound presentation was done through audiophile headphones Sennheiser HD 600.

5.5. Procedure

In each test performance all possible pairs (10 comparisons for the 5 sounds) were presented to each jury using the method of paired comparisons [24, 25] in order to select the one with the highest roughness (the option of equal roughness was not allowed). In order to check the repetitiveness of the jury, two reverse order comparisons were randomly chosen and added to the test. Furthermore, in order to avoid bias in the test, the resulting 12 comparisons were randomly ordered. As an example, Table II shows the ordered comparison pairs for the test involving 1000 r.p.m. signals.

At the beginning of each test session, judges were trained in the sensation of roughness by the presentation of a set of 6 different sounds. The purpose of the training was only to introduce the concept of roughness to the group of non expert subjects forming the jury panel. This sound set consisted of a pure tone of 1000 Hz modulated in amplitude with a frequency of 70 Hz using modulation indexes of 0%, 20%, 30%, 70%, 90% and 100% respectively. The sounds were sorted from lowest to highest modulation index, so the roughness sensation was clearly perceived.

Finally, the consistency and repetitiveness of each judge have been analyzed for each test, not considering those people whose consistency scores were under 80% [24, 26]. As a consequence, two persons were rejected in test 1 and other two different persons were rejected in test 4.

5.6. Results of Jury Test using 1000 r.p.m. Synthesized Noise

In a paired comparison test a final score is assigned to the signals involved in such a way that signals with highest scores correspond to the ones that jury has perceived having more roughness. Figure 9 shows the signal scores for the 1000 r.p.m. test (modulation index and frequencies have been given in Table I). As the sum of the scores provided by the method of paired comparisons is always zero, some of the scores are positives while others are negatives. From these results, some interesting conclusions can be stated:

a) Comparing signals 1, 2 and 3: these signals correspond to modulation frequencies of 50 Hz with modulation indexes of 0% (no modulation, reference Signal 1), 20% and 40%, respectively. Scores obtained indicate that Signal 2 is perceived much rougher than Signal 3, in contradiction with the increasing modulation index used to synthesize these signals.
b) Comparing signals 1, 4 and 5: these signals correspond to modulation frequencies of 70 Hz with modulation indexes of 0% (no modulation, reference signal 1), 20% and 40% respectively. In this case scores obtained seem to be even more opposite to theory because Signal 5 (highest modulation index) is appreciated as the one with the lowest level of roughness.
c) Comparison between Signals 2 and 4 and between Signals 3 and 5: Signals 2 and 4 have the same modulation index (20%) but differ in their modulation frequency. According to theory, a signal modulated at a frequency of 70 Hz should be perceived rougher than one modulated at 50 Hz, so score of Signal 4 should be higher than score of Signal 2. Similarly, score of signal 5 should be higher than that of Signal 3, but results obtained are just opposite to those expected in theory.
d) Score range obtained for the five signals is very short (from −0.3 to 0.4). In paired comparison tests a short range suggests that all signals involved are perceived very similar with respect to the evaluated feature. In fact some jury members report problems to identify the roughness sensation in the synthesized signals. On the other hand, modulation indexes and frequencies used to generate the noises were different enough to clearly generate differences in calculated roughness.

Therefore, the question of how to explain the disappointing results obtained from this first experiment arises. A reliable approach comes from an analysis in depth of the model proposed in Figure 7. The set of impulses $c(n)$ has a fundamental frequency given by (2), that for a 1000 r.p.m. engine noise is

$$f_0 = \frac{1000}{120} \times 4 = 33.33 \text{ Hz}, \tag{23}$$

whose inverse corresponds to a fundamental period of 30 ms ($N = 1440$ samples for a sampling frequency of 48 kHz). Therefore, an impulse is generated every 30 ms, being its amplitude given by equation (22).

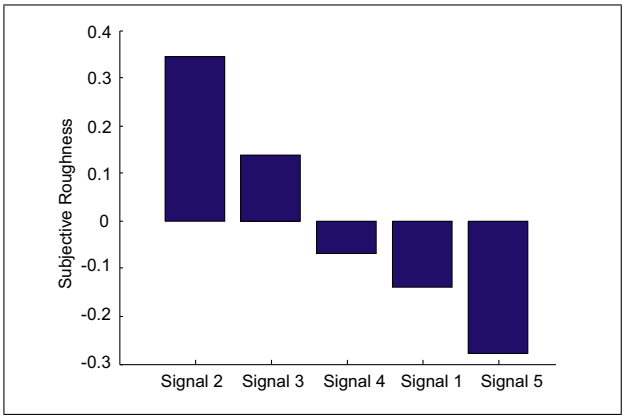


Figure 9. Scores obtained in the 1000 r.p.m. jury test for the set of synthesized signals described in Table I.

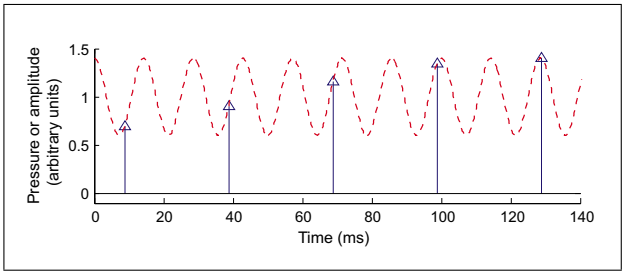


Figure 10. Set of impulses $c(n)$ used in the generation of a synthesized 1000 r.p.m. signal modulated in amplitude with a modulation frequency of 70 Hz and a modulation index of 40 % (Signal 5 in Table I).

This expression shows that the values of a_k come from the sampling of the modulation signal of frequency f_m at a sampling frequency of f_0. For this sampling frequency, Nyquist theorem states that only signals with frequency content up to $f_0/2$ can be sampled properly [27]. Otherwise undersampling makes the sampled signal behave like the alias whose frequency content lies in the interval between 0 and $f_0/2$, being the alias frequencies of f_m,

$$f_m^{\text{alias}} = \pm f_m + l \cdot f_0, \quad \text{with } l = 0, \pm 1, \pm 2, \ldots \tag{24}$$

Returning to our particular 1000 r.p.m. synthetic signal, note that impulse amplitudes a_k have been generated sampling a tone of frequency $f_m = 70$ Hz with sampling frequency $f_0 = 33.33$ Hz through equation (22). Due to the undersampling effect, the resulting modulation frequency is much lower than expected: according to (24), the modulation frequency of 70 Hz has been down converted to an alias of $f_m^{\text{alias}} = f_m - 2f_0 = 3.33$ Hz. Figure 10 illustrates this undersampling effect showing the train of impulses generated by means of (22) with $f_m = 70$ Hz and $f_0 = 33.33$ Hz: the resulting amplitude variations are so small that would fit a tone sampled at the same f_0 but owing a frequency of 3.33 Hz. Similarly, if the modulation frequency were $f_m = 50$ Hz, impulse amplitudes would fit samples of a single tone of frequency $f_m^{\text{alias}} = f_m - f_0 = 16.66$ Hz.

Table III. Modulation indexes I_{mod} (%) and frequencies f_N (Hz) used in equation (22) to generate amplitudes a_k for the five signals involved in $N = 2000$, 3000 and 4000 r.p.m. tests.

Signals	I_{mod}	f_{2000}	f_{3000}	f_{4000}
1	0	0	0	0
2	20	23.33	35	46.66
3	40	23.33	35	46.66
4	20	33.33	50	66.66
5	40	33.33	50	66.66

Therefore, regarding the results of the jury test given in Figure 9, we should state that they do match the theoretical results because Signals 2 and 3 (equivalent to a 16.66 Hz modulation frequency roughness) have been perceived rougher than Signals 4 and 5 (equivalent to a roughness with 3.33 Hz of modulation frequency). Furthermore, the low modulation frequencies actually used to generate synthesized noises also explain the short score range of Figure 9. The problem is that the fundamental frequency $f_0 = 33.33$ Hz of $c(n)$, imposed by the engine velocity, makes it impossible to use modulation frequencies over 18 Hz, which are too low to provide significant differences in roughness perception when tested. For this reason, results obtained from the 1000 r.p.m. test have been discarded in the overall conclusions.

5.7. Results of Jury Tests using 2000 r.pm., 3000 r.p.m. and 4000 r.p.m. Synthesized Noises

As shown in section 5.6, the maximum modulation frequency that can be used to synthesize engine noises is equal to half the corresponding engine velocity, so that maximum modulation frequency for 2000 r.p.m. signals is 33.3 Hz, for 3000 r.p.m. signals is 50 Hz and 66.6 Hz for 4000 r.p.m. noises. Therefore, two different modulation frequencies have been chosen for each test to generate the noises: one has been the maximum frequency allowed and the other corresponding to 70 % of the maximum. Table III summarizes the new parameters of the signals involve in the three tests.

Before analyzing the results of the three jury tests, it must be noted that 15 different marks of subjective roughness have been obtained, but their meaning is constrained to the context of the test in which the signal was evaluated, so a signal can only be compared to another one evaluated in the same test (for example, roughness score of 2000 r.p.m. signals is not compared with that of 3000 r.p.m. signals).

Figure 11 shows the results of the three different tests. Figure 11(a) illustrates the dependence of roughness on the modulation index for a modulation frequency of 70% of the maximum allowed for each engine velocity. The results show that an increase of the modulation index produces an almost linear increase of the roughness perceived. The big range between the maximum and minimum value of the roughness (from −4 to 3) suggests the agreement of the jury panel in deciding which sound was the roughest, as well as which one was the softest along the comparisons.

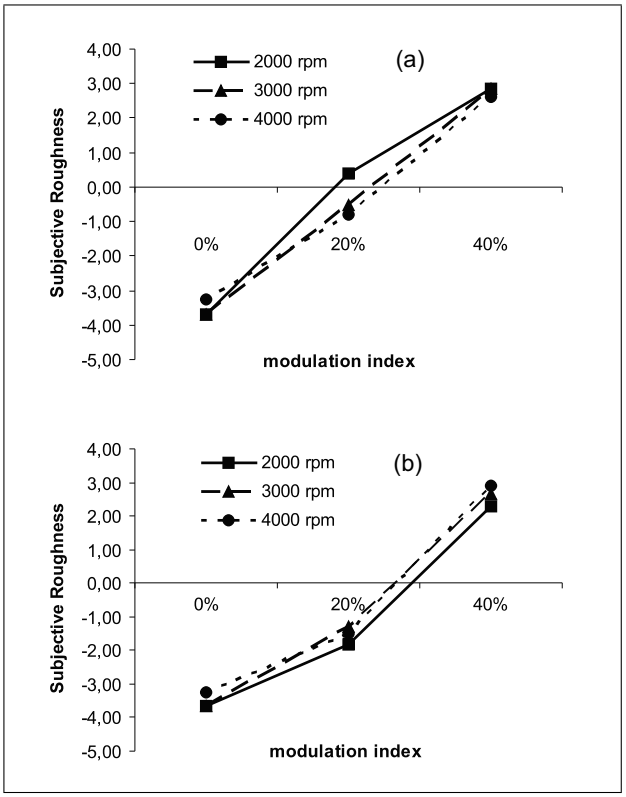


Figure 11. Subjective roughness versus modulation index (0%, 20% and 40%) for 2000, 3000 and 4000 r.p.m. tests. Graph (a) has used the maximum modulation frequency allowed for each engine velocity whereas diagram (b) has used a modulation frequency of 70% of the maximum.

On the other hand, Figure 11(b) represents the dependence of roughness on the modulation index when the modulation frequency is the maximum allowed for each corresponding engine velocity. The results are very similar to those of Figure 11(a), so it can be stated from both graphs that, independently of the modulation frequency used, an increase of the modulation index produces a nearly linear increase of roughness.

With regard to modulation frequency influence, Figure 12 shows the subjective roughness perceived comparing signals with the same modulation index but different modulation frequency (see Table III for actual values of frequency). Figure 12 (a) shows the dependence of roughness on the modulation frequency, when a modulation index of 20% is used. It can be seen that this dependence is somewhat stronger for 2000 r.p.m. sounds than for higher engine velocities. However, all the tests show an slight increase of roughness perceived with respect to the increase on the modulation frequency. The low marks obtained by these signals with respect to those of Figure 12(b) can be explained by the small modulation index chosen (20%).

Finally, Figure 12 (b) represents the roughness of the modulated signal with a 40% modulation index. Its behavior is very similar to that of 20 % except for 4000 r.p.m. noise. In this particular case an increase on the modulation

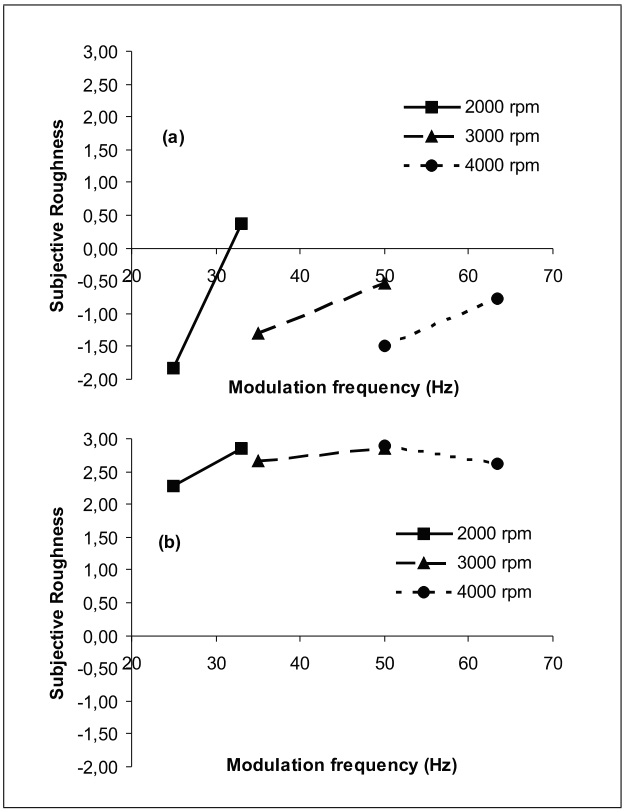


Figure 12. Subjective roughness versus modulation frequency (see Table III for its exact values), with: (a) modulation index of 20%, and (b) modulation index of 40%.

frequency corresponds to a decrease on the roughness perceived. Nevertheless, it can be appreciated from the Figure that roughness marks are very similar, indicating that dependence on the modulation frequency is not so important as on the modulation index.

Although the jury tests carried out for synthesized noises do not pretend to be exhaustive in the number of engine velocities, modulation frequencies and modulation indexes involved, their results are quite representative of the behavior of the model proposed in Figure 7. Results of Figure 11 confirm that model based on cepstrum analysis is able to generate synthesized noises with controlled roughness. Moreover, the model proposed identify the modulation index in equation (22) as the parameter with the strongest influence over the roughness perceived.

In summary, two important conclusions can be extracted from Figures 11 and 12. Firstly, the model proposed in Figure 7 has been validated as a good method to synthesize car engine noise under controlled roughness conditions. Moreover, this result allows to consider parameters of the complex cepstrum analysis for the improvement of roughness calculation methods when applied to broadband noise signals, instead of use only parameters extracted from narrowband models. The second important conclusion is that subjective roughness has a dependence on the modulation index quite stronger than on the modulation frequency (by comparison of the different ranges covered by scores on Figures 11 and 12).

6. Conclusions

Roughness sensation for car engine noises has been explored through complex cepstrum analysis of the real signals. Once the complex cepstrum model has been obtained, a synthetic waveform generator has been developed to simulate noises in order to relate model parameters to roughness perception. The synthesized engine noises are composed by two basic signals added together: a colored noise and a signal produced by the repetition of a pattern extracted from a real noise through complex cepstrum analysis. The model can produce sounds with controlled roughness using different pattern modulation depths and modulation indexes. Roughness of such generated noises presents a high dependence on the modulating index and a weak dependence on the modulation frequency when amplitudes of the repeated patterns are calculated through an Amplitude Modulated (AM) signal model.

Acknowledgement

This work has been partially supported by Spanish Government and Regional Government (Generalitat Valenciana) through grants TIC2003-08238-C02 and GV04A-444 respectively.

References

[1] ISO 532-1975: Acoustique méthode de calcul du niveau d'isosonie.

[2] E. Zwicker, H. Fastl: Psychoacoustics, facts and models. 2nd. ed. Springer-Verlag, 1999.

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

[4] W. Aures: Ein Berechnungsverfahren der Rauhigkeit. Acustica **58** (1985) 268–281.

[5] B. C. J. Moore: An introduction to the psycology of hearing. 5th ed. Academic Press, 2003.

[6] J. Blauert: Spatial hearing: the psychophysics of human sound localization. MIT Press, 1997.

[7] H. L. F. v. Helmholtz: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. F. Vieweg, Braunschweig, 1863.

[8] E. Terhardt: Ein psychoakustisch begründetes Konzept der musikalischen Konsonanz. Acustica **36** (1976) 121–136.

[9] P. Daniel, R. Weber: Psychoacoustical roughness: Implementation of an optimized model. Acustica united with Acta Acustica **83** (1997) 113–123.

[10] A. Moreno, F. Simón, C. de la Colina, A. Marchioni: Factores determinantes en el calculo de la aspereza de señales acusticas por el metodo de Aures (in Spanish). Proc. of Tecniacústica in CD-ROM, Avila, 1999, 1–7.

[11] B. Feng, G. Wakefield, M. Good, N. Otto, R. Rebandt: On the perceived roughness of automotive powertrain sounds. SAE paper 951312, 1995. 639-647.

[12] M. Pflueger, R. Hoeldrich, F. K. Brandl, W. Biermayer: Subjective assessment of roughness as a basis for objective vehicle interior noise quality evaluation. SAE paper 1999-01-1850, 1999.

[13] A. Gonzalez, M. Ferrer-M, M. De Diego, G. Pinero, J. J. Garcia-Bonito: Sound quality of low-frequency and car engine noises after active noise control. Journal of Sound and Vibration **265** (2003) 663–679.

[14] B. P. Bogert, M. J. R. Healy, J. W. Tukey: The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. – In: Time Series Analysis, ch. 15. M. Rosenblatt (ed.). 1963, 209–243.

[15] A. V. Oppenheim, R. W. Schafer: From frequency to quefrency: A history of the cepstrum. IEEE Sig. Proc. Magaz. **21** (2004) 95–106.

[16] K. Schmillen, J. Wolschendorf: Cycle-to-cycle variations of combustion noise in diesel engines. SAE Paper 890126, 1989.

[17] S. Cerdá, J. Romero, J. Navasquillo: Análisis tiempo-frecuencia de las resonancias en la combustión (in Spanish). Revista de Acústica **32** (2001).

[18] T. Patro: Combustion induced powertrain NVH. A time-frequency analysis. SAE Paper 971874, 1997.

[19] A. V. Oppenheim, R. W. Schafer: Discrete-time signal processing. 2nd ed. Prentice Hall, 1999.

[20] R. Hoeldrich, M. Pflueger: A generalized psychoacoustical model of modulation parameters (roughness) for objective vehicle noise quality evaluation. SAE paper 1999-01-1817, 1999.

[21] B. Feng, N. Otto: Synthesis of powertrain sounds for investigations in roughness. SAE Paper 931333, 1993. 571-578.

[22] L. B. Jackson: Digital filters and signal processing. 2nd ed. Kluwer Academic Publishers, 1989.

[23] M. Ferrer, M. de Diego, G. P. nero, A. González: Evaluación subjetiva de la calidad del sonido mediante test de parejas (in Spanish). Proc. of Tecniacústica in CD-ROM, Valladolid, 1999, 1–7.

[24] H. A. David: The method of paired comparisons. Oxford University Press, New York, 1988.

[25] R. A. Bradley, M. A. Terry: Rank analysis of incomplete block designs. 1. The method of paired comparisons. Biometrika **39** (1952) 324–345.

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

[27] A. V. Oppenheim, A. S. Willsky, S. Hamid Nawab: Signals and systems. 2nd ed. Prentice Hall, 1997.