Development of a metric to quantify diesel engine irregularities

G. Lowet, P. Van de Ponseele, *S. Pauwels, *T. Van Wayenberge and *P. Sas
LMS International, Interleuvenlaan 68, 3001 Heverlee, Belgium
*Department of Mechanical Engineering, division PMA, K.U.Leuven, Belgium
e-mail : geert.lowet@lms.be

Abstract

This paper describes the development of a metric to quantify the perception of irregularity in diesel engine sound. This development started from the formulation of hypotheses on signal parameters that possibly had an influence on the perception. Using dedicated signals, the validity of each of the hypotheses was tested using a listeners test. Based on the results of the tests, a metric was developed that took all of the important signal parameters into account. Listeners tests were also made evaluating unmodified recordings of diesel engine noises, using four cylinder engines in cold idle condition. A linear correlation of $r^2=0.80$ between the jury test results and the metric calculations was found, indicating that the metric accounted properly for the main sound parameters determining the preception of irregularity

1. Introduction

In the context of diesel engine sound quality in stationary conditions, mainly attention is paid to the impulsive character of the sound. This effect is related to sudden changes in the signal amplitude. Another characteristic is the irregularity of the signal. The irregularity of the sound is to be differentiated from its impulsiveness. While impulsiveness is mainly related to changes in the signal envelope, and can be very regular in time, irregularity can be described as sudden, short-time changes in signal timbre, sometimes accompanied by a change in signal level, and occurring randomly over time.

Whereas impulsiveness is related to the general sound comfort, irregularities in a sound attract the listener's attention and may result in a negative product quality perception. Several attempts have already been made to quantify the diesel engine noise character, mainly by Russell and co-workers [2, 3].

In this paper the methodology, the implementation and the results of the development of a metric for diesel engine irregularities is presented, resulting in a sound metric that is representative for the listeners' response to this type of sound.

2. Methods

In a first step, a set of engine recordings were intensively listened to and analysed. It was observed that, depending on which cylinder is at the origin of the irregular event, irregularities can have a very different timbre: for the 4-cylinder engines, 4 different types of irregularities could be distinguished. Using filtering techniques, it appeared also that irregularities are mostly in the higher frequency band (above 1000 Hz) while the frequency content of the regular part of the signal is mainly below 1000 Hz. This property allows to separate the regular part of the signal from the irregularities.

In a next step, a list of signal parameters was established, both time- and frequency-domain parameters, that were assumed to influence the perception of irregularities in the sound. This was based on known psycho-acoustic effects, and on own impressions after listening to the recordings. For each of these parameters, a hypothesis was formulated, and evaluated by listening tests. The following hypotheses were investigated:

a) Increased sound pressure levels of the irregularity as compared to the overall signal level, will result in an increased perception of irregularity
b) Increasing the total sound level will result in a decreased perception of irregularity

c) Signals where the irregularities have the same character are perceived as less irregular than sounds in which irregularities with different characters are present
d) Signals where the irregular events occur at a fixed intervals are perceived as less irregular than signals where the irregular events occur randomly over time
e) Signals with more irregular events are perceived as more irregular than signals with less irregular events
f) Signals from an engine at high speed are perceived as less irregular than a similar sound resulting from an engine at lower speed

For each of the hypotheses, a set of test sounds was created. These were obtained by editing two real diesel engine recordings in LMS CADA-X Sound Quality Monitor, such that all parameters of the sound were kept unchanged, except the one that was the subject of the hypothesis. The signals were then evaluated for perceived irregularity in a Jury Test. The results of these tests were used to validate or falsify all of the listed starting hypotheses and to define which sound characteristics should be implemented in the metric. The metric was developed such that it corresponded maximally to the results of these tests.

For the *hypothesis (a),* the high-frequency part of the signal was isolated, its level adapted (-4, -2, +0, +2 and +4 dB), and added to the background signal. For *hypothesis (b),* the level of the signal was set at 5 different SPL levels, ranging between 58.5 to 70.5 dB. The signals used for *hypothesis (c)* were obtained by isolating a regular cycle and four different irregular cycles (each relating to an irregularity caused by a different cylinder) from an engine recording. New signals were synthesised then by combining regular and irregular cycles. A first signal contained only one type of irregularity, while in the others gradually more variation in the types of irregularity was introduced. The same cycle segments were used for the signals for *hypothesis (d).* In this case, only one type of irregularity was used. While the first signal contained a constant number of regular cycles in between two irregular cycles, increasingly more variation was allowed in the number of regular cycles in between the irregularities for the other signals. The averaged number of regular cycles was kept constant however. For *hypothesis (e)* a completely regular sound was taken as a start signal. Subsequently 20, 40, 60 and 80% of the regular cycles was replaced by irregular cycles. The effect of engine speed -*hypothesis (f)*- was simulated by playing back a same sound at different speeds, which results in a perception of higher engine speed. The eventual effect of the frequency shift on the listener's response was neglected.

Also 10 unmodified recordings of diesel engine sounds were evaluated in the jury test. The recordings were made on 4 cylinder engines in cold idle condition. In order to reduce the test time, two sets of 6 recordings were made, 2 engine sounds being common to both sets such that the results for the individual sets could be combined for later analysis. The results of these tests were used to validate the metric, and to allow some fine-tuning.

Voluntary non-expert subjects, 100 in total, participated to the listening test. An A/B comparison test with replication was used. Participants were asked to identify the most irregular sound of the presented pair. Because of the size of the complete data set, not all sound sets were presented to all test persons: each test person had to quote two edited sound sets, and one set of real car sounds. By consequence each sound set used for hypothesis testing was judged by about 30 subjects. Each car sound set was judged by half of the group.

The jury test included also a general hearing test and a limited training phase, in order to familiarise the subjects to the type of sounds and to clarify the concept of irregularity.

3. Results

3.1 Hypothesis evaluation

A preference matrix was made for each participant, and for each test, representing the subject's preference of one sound over another. In fact two matrices were constructed, one for the A-B and one for the B-A comparisons. The consistency of the data for a single subject was evaluated by a number of methods such as the coefficient of consistency, the rank correlation coefficient and the coefficient of concordance. Also the agreement of the subject with the total group was evaluated.

In general low individual consistencies were found for most of the tests. This is largely due to the difficulty of the test, as appeared from the feedback given by the participants after the test. Nevertheless, when the results of all participants were pooled, the overall result allowed validating or invalidating the formulated hypotheses. In the figures that illustrate the results, the score indicates the total number of times that a given sound was selected as being the most irregular in the paired comparison. The results for the hypotheses tested are the following:

a) The level of the irregularities with respect to the overall noise level was found to be important. Signals with increased level of the irregularities were found to be more irregular as is clearly illustrated in Figure 1.

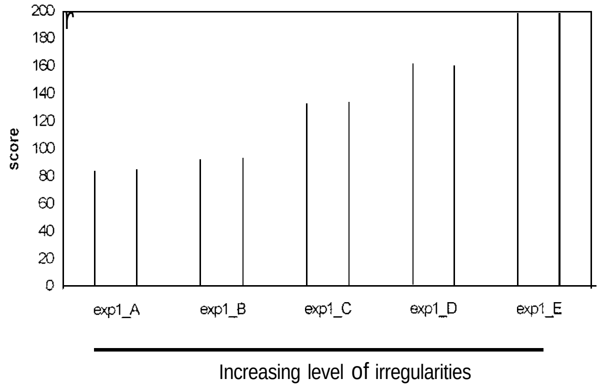

Figure 1: Influence of the level of irregularities on perceived irregularity

b) Increasing the overall level sound resulted in a subjective evaluation of increased irregularity, which is opposed to the hypothesis formulated. The rationale behind the hypothesis was that higher overall sound levels would lead to an increased masking of the irregular components. The discrepancy between the hypothesis and the results might be due to a change in decision criterium of the subjects in this test. More specifically it is suspected that, since no other sound characteristic except its level had been modified, that most subjects -involuntary- have quoted for loudness -or for general comfort- rather than for irregularity. The results are displayed in Figure 2.

c) Sounds synthesised with only a single type of irregularity were judged to be less irregular than sounds containing different types of irregularities, as can be seen on Figure 3. However, once the signal contains different types of irregularities, adding more variation in the irregular types is not perceived anymore. As in all of the recordings made on real engines, the situation of only a single type of irregularity was never met, such situation was considered as very unusual. Therefor this aspect was not considered in the further development of the metric.

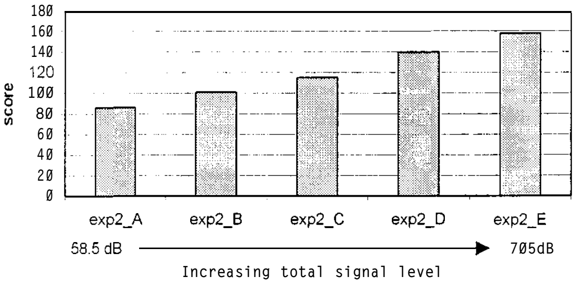

Figure 2: Influence of total signal level on perceived irregularity

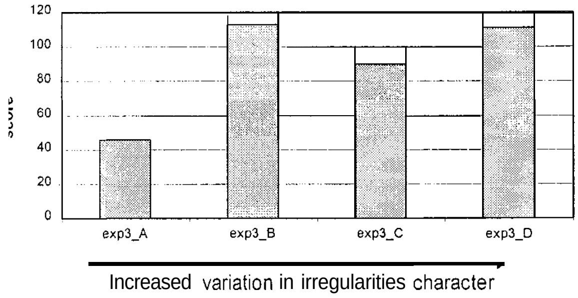

Figure 3: Influence of increased variation in the types of irregularity on perceived irregularity

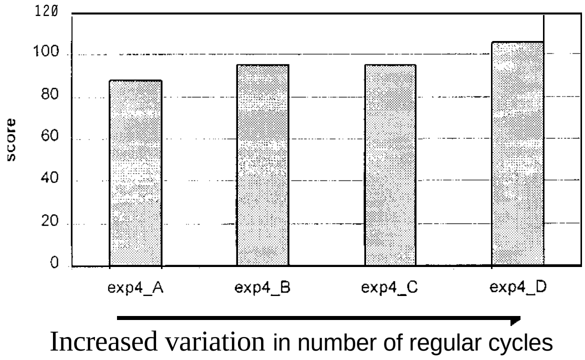

Figure 4: Influence of variation on the number of regular cycles in between irregularities on perceived irregularity

d) Figure 4 shows that signals with a fixed number of regular cycles in between irregular events were not found to be significantly different from sounds where the number of regular cycles was randomly varied. It appears thus that regularities that happen at constant intervals (comparable to impulsive events) are not perceived less irregular than irregularities that happen randomly over time.

e) It is clear from Figure 5 that an increased occurrence of irregularities results in the perception of increased irregularity.

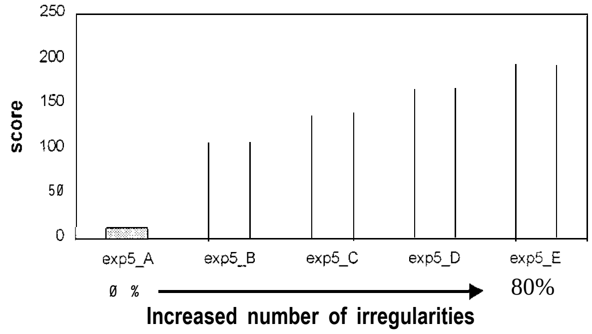

Figure 5: Influence of number of irregularities on perceived irregularity

f) Based on the effect of temporal masking, by which short events can be masked by preceding sounds, the hypothesis was that similar sounds would be perceived less irregular as the engine speed increased. Higher idle engine speeds result in a lower perceived irregularity but it is limited to 900 rpm: higher rpm signals are not perceived as more irregular (Figure 6).

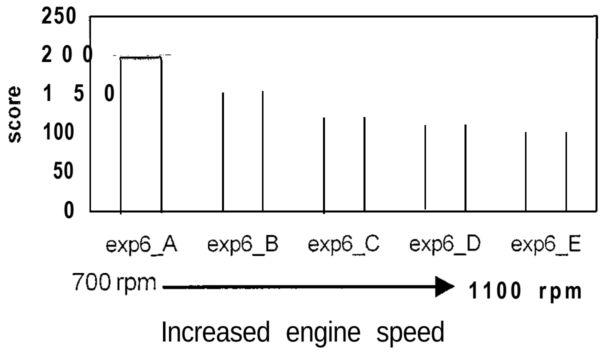

Figure 6: Influence of engine speed on perceived irregularity

As a result from these tests, it was concluded that the metric to be developed needed to account mainly for the level of the irregularities, the overall signal level, the number of the irregularities and the engine speed.

3.2 Definition of the irregularity metric

A prototype metric was developed, based on the work of Lelong [1]. Its is however based on the A-weighted Energy content within a cycle of the signal. The formulation found is:

$$Irreg = \sigma\{E_A(i)\}.Max^2(T,0.033).\overline{E_A(i)} \qquad (1)$$

T is the duration of one ignition cycle, and

$$E_A(i) = 10\log\left[\int_{t_i}^{t_i+1} \frac{p_A^2}{p_0^2}\,dt\right] \qquad (2)$$

In this formula p_A is the A-weighted pressure, p_0 the reference pressure. Using the A-weighting for the sound pressure emphasises the higher frequencies and, by consequence, gives relatively more weight to the irregularities in the signal. The integration in formula (2) is made over a total ignition cycle, such that EA in fact represents the total acoustic energy within that cycle. The saliency of the irregularities is expressed by calculating the standard deviation of EA, while the overall level is introduced in the metric by the mean value of EA. The effect of engine speed is reflected in a multiplication by T, the mean cycle duration, however limited to 0.033 seconds, to be in agreement with the listening test results.

3.2.1 Correlation of the metric with jury test results

In order to compare the jury test results to the metric calculation, the Bradley-Terry model (See ref. [4]) was used to transform the results of the jury test to a linear scale, in which the results are expressed by their *merits*. The correlation of the metric and the merits calculated from the jury tests on the signal sets used for the hypothesis testing was very high with r2 values ranging between 0.957 and 0.991. This indicates that all effects observed in the jury tests are adequately covered by the metric formulation.

Also the jury tests results obtained from the ten unmodified engine recordings were transformed to a linear scale, such that the correlation between the merits and the metrics result could be evaluated. The result is displayed in Figure 7. A highly significant correlation between the jury test results and the calculated metric was found ($r^2=0.795$, $p<0.001$).

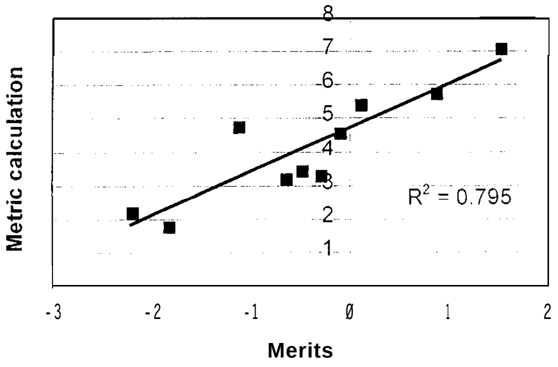

Figure 7: Relation between measured merits and calculated irregularity for 10 engine sounds

4. Discussion and Conclusion

The methodology used in this development is based on the investigation of isolated and well-defined signal characteristics. This approach was preferred over a purely mathematical model, relating the jury test results to a combination of standard psycho-acoustic parameters, such as loudness, sharpness etc, or spectral and temporal signal characteristics. It guarantees that the metric properly accounts for the basic signal characteristics that contribute to the perception of irregularity, and reflects a causal relationship, contrary to a statistical model.

The disadvantage is that the success of the method depends on the relevance and completeness of the hypotheses formulated. If the hypotheses do not describe all of the important elementary aspects of the perception of a phenomenon, the method might lead to a metric that fits each of the hypotheses, but does not adequately quantify complex sounds. This would be reflected into a poor correlation between the metric and the evaluation of real engine sounds, and could result in a need to formulate and test additional hypotheses.

The jury test method used, A/B with replication, was found to be satisfying for the participants to this test. The fact that a preference had to be given (equality between sounds was not allowed), and the fact that differences between sounds were sometimes quite subtle, explain the rather low consistency observed within the individuals. Nevertheless the pooled data were very clear with respect to the hypotheses made. It seems therefore that these inconsistencies can be considered as “random noise” for the high number of participants involved.

It was concluded that mainly the level of the irregularities, the signal level, the number of irregular events and the engine speed were important for the perception of irregularity. The metric developed took into account each of the individual aspects that were found to be relevant for the perception of irregularity. When applied to the real engine sounds, it was found that the metric was in good agreement with the perception of irregularity as measured in the jury test. About 80% of the variance on the listeners’ evaluation is covered by the metric. Because of the complexity of the perception, and the uncertainty explicitly expressed by many of the test persons, it is the opinion of the authors that the 20 % unexplained variance might be mainly an expression of the subject inconsistency, and the random “guessing” which adds a part of unexplained variance, and less due to incompleteness of the hypotheses set.

Considering the limitations of the present study, it can be concluded that the present approach led to a psycho-acoustic based development of a sound metric. This metric has been proven to be representative for the subjective perception of the irregularity of 4-cylinder diesel engine noise.

References

1. Lelong *Estimation of road traffic noise using Laeq measurements*, Internoise 97, pp.899 (1997)
2. M.F. Russell, M. Sekowski, N. Nikokiroulis *Noise quality in small diesel engined vehicles,* 2nd Int. Conference on Vehicle Comfort, Bologna, Italy, Technical papers, Vol. 2, pp.985-99 (1992)
3. M.F. Russell, M. Sekowski, N. Nikokiroulis *Subjective assessment of diesel vehicle noise,* IMechE, FISITA 92, Technical paper C389/044, pp.37-42 (1992)
4. H.A. David *The method of paired comparison.*,Ed. Charles Griffin & Company Ltd, Oxford University Press, New York (1988)