Just-Noticeable Roughness Differences of Technical Sounds

A. Leutheuser[1], R. Sottek[2], J. Becker-Schweitzer[1]
[1] *Fachhochschule Düsseldorf, Institut für Schwingungstechnik, Josef-Gockeln-Str. 9, 40474 Düsseldorf, Germany*
Email: alexandra.leutheuser@gmx.net
[2] *HEAD acoustics GmbH, Ebertstr. 30 a, 52134 Herzogenrath, Germany*
Email: roland.sottek@head-acoustics.de

Introduction

Roughness has become a central focus for product sound design. The roughness-induced perception, varying from a sporty character to a very unpleasant impression, proposes consistently new questions and challenges in sound engineering. Just-noticeable roughness differences of synthetic and engine sounds have been studied considering the influence of a single signal parameter as well as a combination of parameters. The results of these studies will be used to improve the roughness calculation based on the Hearing Model according to Sottek with respect to technical sounds.

Roughness sensation

The roughness sensation is caused by envelope variations with a modulation rate between 15 and 300 Hz. The unit of roughness is asper. A sine tone of 1 kHz with a level of 60 dB, 100% amplitude-modulated at a rate of 70 Hz is defined to have a roughness of 1 asper. The important parameters are degree of modulation, carrier frequency and modulation rate. The signal level has only a small influence on the roughness impression [1]. In the case of slow envelope variations (below 15 Hz), the human ear is capable of tracking the loudness changes, resulting in an impression of fluctuation. With increasing modulation rate other sound impressions are perceived, such as engine roughness (for rates between 20-40 Hz), which then changes to the psychoacoustic roughness impression.

Just-noticeable roughness differences

The **J**ust-**N**oticeable **D**ifference (JND) describes the smallest detectable difference between two stimuli. The studies by Fastl and Zwicker have revealed that in the case of amplitude-modulated pure tones an increment of roughness becomes perceptible if the degree of modulation is increased by 10%. The corresponding increment in roughness is about 17% [2]. Almost all investigations have been confined to experiments with amplitude-modulated pure tones with varying degree of modulation. The published thresholds related to the degree of modulation have been in the range from 5% [3] to 25% [4] of relative changes. This substantial discrepancy can be explained by different parameters of the stimuli. E.g. the sound pressure level in [3] was about 20 dB higher than in [4].

This paper concerns further experiments performed to investigate just-noticeable roughness differences of synthetic and technical sounds. The primary aim of these studies is to obtain information about the tolerances in the calculation of roughness, which will be used for optimizing the Hearing Model according to Sottek [5].

Methods of JND determination

For the determination of just-noticeable differences, two different approaches are often used: procedures of classical psychophysics and adaptive methods.

The method of adjustment, the method of limits and the procedure of constant stimuli belong to the classical methods. They are all based on comparing one reference sound to other sounds of a test set having different parameter values; only the procedure of adapting the stimulus strength is different. The disadvantages of these methods are on the one hand the high time expenditure and on the other hand the dependency of sensory evaluation on individual characteristics of test persons such as assiduousness while performing the listening tests [6]. In order to avoid this problem and to improve the measurement efficiency additional adaptive procedures have been developed, where the value of the stimulus strength is adapted according to the subject’s previous responses. These methods are combined with a paradigm e.g. in the form of a yes-no-method or some forced-choice methods. The forced-choice methods (AFC) are based on a selection procedure and involve two (2AFC) or more sounds (3AFC, 4AFC) presented one after another; only one of these sounds differs with respect to a certain attribute. The subject’s task is to indicate which of the sounds is different.

How adaptive procedures work will be explained in the following using the examples of the staircase method (up-down and transformed up-down) and the Best-PEST-procedure.

The simple up-down procedure proposes an adaptive rule, according to which perceived differences between reference and target stimulus (correct subject’s answers) lead to stimulus strength reduction and not perceived differences (incorrect answers) to an increase. The measurement results correspond to a target stimulus in which the tracks of the answers level off around the threshold. A drawback of this procedure is the simple rule. Subjects could guess it and could manipulate the threshold results.

The transformed up-down procedure represents a first optimization because more complicated adaptive rules are used, for example the one-up-two-down rule: in this case stimulus strength is reduced after two subsequent correct answers and increased after every incorrect answer. Its advantage is the elimination of manipulation risk, but quite a high measurement time is needed.

Therefore another more efficient and faster procedure has been used for the following investigations: the Best-PEST-method developed in 1980 by Alex Pentland [7]. It is based on the PEST-method (***P**arameter **E**stimation by **S**equential **T**esting*) worked out by Taylor und Creeman [8] and it is one of the first methods which consider not only directional changes of stimulus strength, but also the step size.

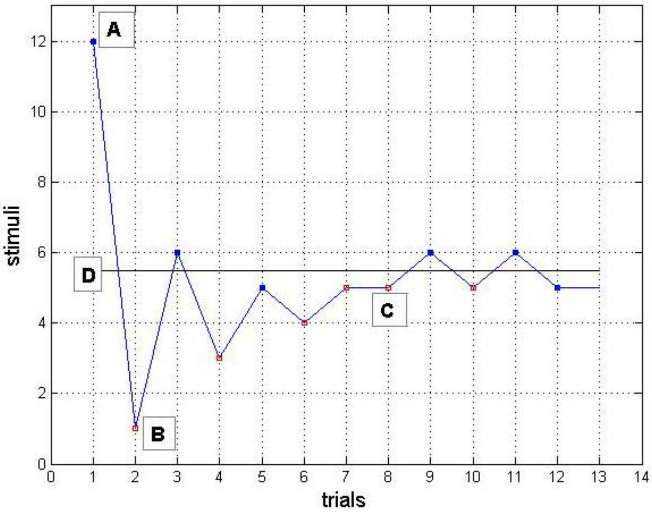

Figure 1: Adaptive tracks according to a Best-PEST-procedure as a function of the range of stimuli and the resulting trails. Correct responses lead to a downward move, incorrect answers lead to an upward move. The reference sound corresponds to the lowest stimulus index (0).

Figure 1 demonstrates adaptive tracks according to a Best-PEST-procedure. The ordinate shows the range of stimulus strength with its increasing intensity (reference sound corresponds to the lowest stimulus index (0)) and the abscissa shows the trials. Just as with the mentioned procedures, BEST-PEST is based on comparing a reference signal to other signals from a test set. In the same way as before, correct answers (Figure 1, A) cause stimulus strength reduction and incorrect answers (Figure 1, B) cause increase. The necessity of stimulus strength change is determined with the help of a maximum-likelihood method. This method calculates the likelihood for every stimulus being the threshold, taking into account all of the previous subjective responses, not just the last one. If the same stimulus gets different evaluations the trial should be performed again (Figure 1, C). This procedure begins with a stimulus in the middle of the stimulus range and then in the following the intervals are divided by a factor of two after each change of direction. When the smallest possible interval (step size) of stimulus range is reached, it remains there towards the end of the measurement (maximal number of trials is carried out or the abort criterion according to the maximum likelihood method is fulfilled). The threshold as the result of the measurement (Figure 1, D) can be found in the corresponding interval of the stimulus strength. For further detailed information about adaptive methods please refer to [7], [8], [9] and [10].

Investigations with synthetic sounds

The purpose of the investigations with synthetic sounds was to study the perception of relative changes of modulation. By means of adding noise, possible artifacts in modulation and carrier frequency should be masked and therefore the test subjects should better focus on the roughness sensation.

It should be investigated to what extent masking effects according to different **S**ignal-to-**N**oise-**R**atios (SNR) would occur. For this purpose amplitude-modulated pure tones with additional pink noise were generated and the degree of modulation and the SNR were varied. As carrier frequency, 1 kHz was chosen, with a modulation rate of 70 Hz and level L=60 dB(A). The degree of modulation m was varied within the range of 0.4-0.76, 0.5-0.86 and 0.6-0.96, depending on the three reference signals with m=0.4, 0.5 and 0.6. This leads to 9 different experiments in combination with three different SNRs (15, 24 and 30 dB). The thresholds were determined via the mentioned Best-PEST-procedure using a kind of „unforced-choice“-method [11]. A 3AFC-procedure was combined with a binary yes/no task. Subjects had the alternative to answer “all sounds are equal” in case of inaudible differences. Thus the guessing rate could be reduced and the efficiency increased. The trials occurred twice in random sequence, while the repetition had the same order of sequence. The subjects had to determine the roughest sound out of the three stimuli presented.

Before every listening test the subjects could listen to three anchor sounds covering the whole range of stimuli. So the participants got an idea of the stimuli to be judged. Further there was a training phase, where the subjects were instructed how to use the software interface.

The first investigations with a SNR of 30 dB and 15 dB were performed with eight subjects (one female, seven male, average age 29 years). Fifteen subjects (two female, 13 male, average age 33 years) have participated in the tests with a SNR of 24 dB.

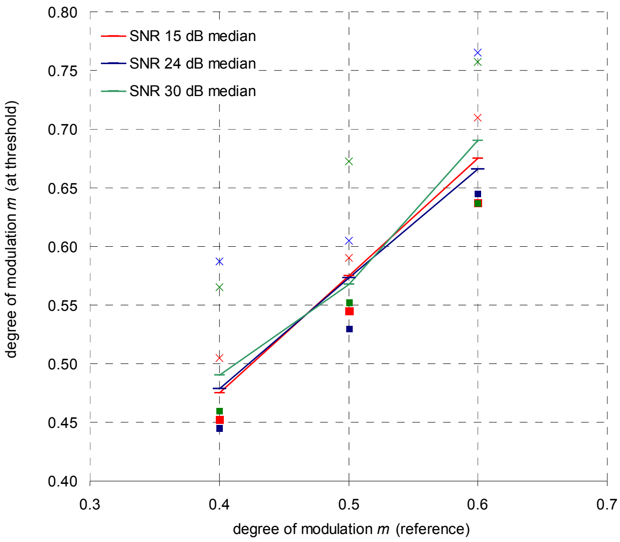

Figure 2: Degree of modulation m of a 1 kHz-tone modulated with 70 Hz required for an audible roughness difference, shown for three different reference signals (degree of modulation: 0.4, 0.5 and 0.6) and three SNR-values (15 dB: red, 24 dB: blue and 30 dB: green). The lines correspond to the median values. Maxima are represented by crosses and minima by squares.

Figure 2 depicts the results: the thresholds are taken as median values (connected with solid lines). The thresholds related to the degree of modulation m=0.5 do not depend on

SNR and are about 13.5-14% (relative change). For the reference sounds with the degrees of modulation *m*=0.4 and 0.6 an increased variance can be seen. The signals with a SNR of 30 dB show basically higher thresholds (22.5% for *m*=0.4, and 15% for *m*=0.6), though the results of tests with a SNR of 15 dB prove lower variance.

The listening test results demonstrate that additional noise regardless of SNR does not influence the thresholds of roughness differences. The determined thresholds are for a degree of modulation *m*=0.4 in the range from 17% (for SNR=15 dB and SNR=24 dB) to 22% (for SNR=30 dB), for *m*=0.5 about 12% and for *m*=0.6 between 10% (for SNR=24 dB) and 15% for (SNR=30 dB).

Investigations with technical sounds

For the second investigation engine sounds were used. By means of these listening tests the threshold of audible differences related to different modulation spectra was studied, especially different contributions at the modulation rates 30 Hz (stronger influence on engine roughness) and 60 Hz (more related to psychoacoustic roughness).

Pretests have shown the requirement of having very similar sounds differing only in roughness, in order to obtain stable results without enormous variances. First tests generating different rough sounds based on measurements by means of order- tracking filters were not successful overall. The control over the roughness of the generated sounds was not precise enough. Finally one engine sound was modified targeting an increased or decreased roughness by means of a binaural transfer path analysis and synthesis. Therefore all relevant transfer paths from a combustion engine (airborne and structure-borne excitation signals) to the driver's ears in the passenger compartment were determined in a first step using the Binaural Transfer Path Analysis (BTPA). Binaural Transfer Path Synthesis (BTPS) is the process of creating listenable vehicle interior noise data based on a BTPA model or modifications of it [12].

Each individual path can be auditioned independently, to assess their respective impact on the overall sound quality. The sum of all contributions agreed well with a reference measurement inside the vehicle interior. The three paths contributing mainly to the perceived roughness were identified. Then a new synthesis model was set up using different weighting factors in the range of 0-2.85 for these paths, thus generating similar sounds evoking different roughness sensations. It was intended to increase the contribution of these rough components. A signal segment of 1.5 s was taken from an engine run-up to achieve a quasi-stationary character. The overall loudness of all stimuli was kept constant (17 sone).

In that way stimuli for two tests were generated. Each test consisted of 20 sounds. For the first test the signals were amplified by +6 dB around 300 Hz, the frequency range with the strongest modulation, in order to sensitize the subjects on the roughness differences in a training phase. Figure 3 shows the modulation spectra of the anchor sounds from the first (left) and the second test (right). The signal named “sound_1” evokes the less rough (without the strong modulated paths), “sound_10” the medium rough and “sound_20” the roughest sensation. The sounds from the first test show higher modulation at 30 Hz, whereas the signals from the second test show almost equal modulation at 30 and 60 Hz.

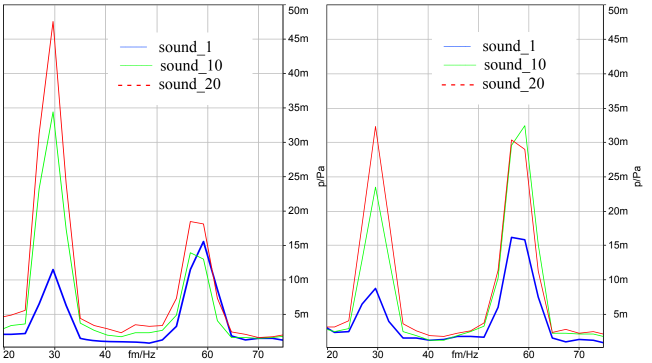

Figure 3: Modulation spectra of anchor sounds from the first test (left) and second test (right).

Then, the sounds were ranked according their roughness using paired-comparison tests. A short pretest gave evidence of a threshold within the first half of the range of stimuli. Thus, the time for the listening tests could be reduced and the subjects had to compare 11 engine sounds only, whether one was rougher than the other or not. For each comparison the roughness index of the rougher sound was incremented by 1 or remained unchanged if no roughness differences were detectable. Thus, comparing 11 sounds, a maximal index of 10 would be possible. Because in some cases no roughness difference was audible, the maximal index was lower (Figure 4).

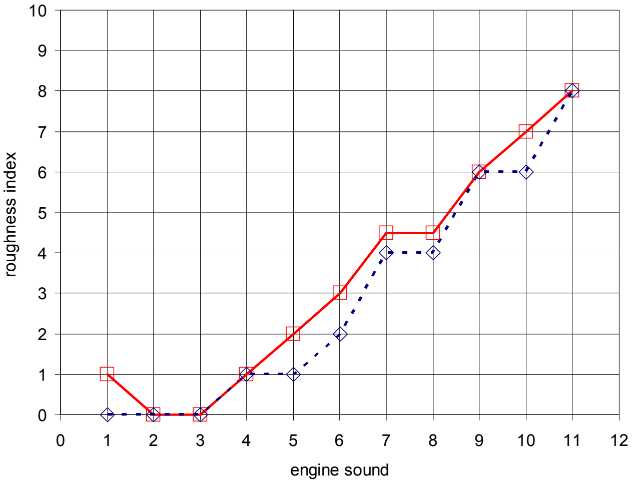

Figure 4: Roughness index resulting from paired-comparison tests with 11 engine sounds for two different test sets based on the same BTPS results. The median values of 11 individuals are shown.

For this listening test, 11 test subjects were used (mainly experts, one female, ten male, average age 28 years). Results of both investigations are represented by means of median values in Figure 4. For the following determination of just-noticeable roughness differences, the sequence of sounds was adapted according to their roughness indices, i.e. sound 1 from the subset represented by a solid line (test 1) in Figure 4 was placed between sounds 3 and 4. This was not necessary for the second test (dashed line).

The thresholds were determined as in the experiment with the synthetic signals, using an adaptive forced-choice procedure comparing e.g. all engine sounds to the engine sound with the lowest roughness index. In these listening tests 30 test subjects took part (3 female, 27 male, average age 32 years, the youngest proband was 24 and the oldest was 61 years old).

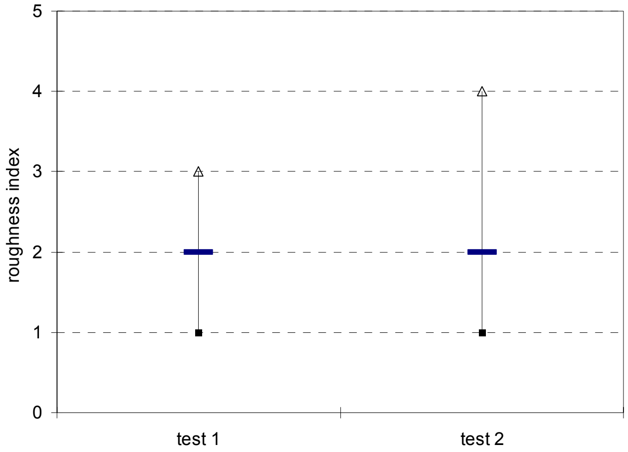

Figure 5: Engine sound required for an audible roughness difference: the reference sound is the sound with the lowest roughness index (Figure 4). Minimum, median and maximum values for two experiments using different combination of the strong modulated paths are shown. The second test points to a slightly higher variation of threshold.

The results of the experiments are depicted in Figure 5. The reference is the sound with a roughness index of zero. Test 1 used as training with filtered sounds points to a threshold corresponding to a roughness index 2 (Figure 4). The results of the second test show higher variance compared to the first test which might be explained by the influence of the stronger 60 Hz modulation (psychoacoustic roughness).

Optimizing the Hearing Model

The results of the threshold experiments are being used to optimize the Hearing Model according to Sottek. On the one hand they delivered information about the necessary accuracy of the roughness calculation; on the other hand the ranking of the stimuli can be used for further studies correlating measured subjective data and calculated roughness.

Investigations by [5], [13] and [14] have demonstrated a high correlation between subjective data and calculated roughness based on the Hearing Model, especially compared to other algorithms. However it must be noted that in general the calculated roughness values are rather small for technical sounds in comparison to synthetic sounds. Additional studies for possible reasons indicated that the specific roughness conform quite well to the results of listening tests [12]. Consequently the weighting of the envelope fluctuations have to be improved after taking technical sounds into consideration. The optimization is being carried out at the moment. So far, processing at low carrier frequencies and low modulation rates is changed in a way that the correlation between measurement results and calculated roughness values is even higher, also for amplitude-modulated pure tones.

Optimizing a model requires many listening tests. Up to now similar stimuli were used. In the next step experiments with other technical sounds must follow.

Conclusion

Results of the roughness threshold experiments with amplitude-modulated signals with additional noise components correspond to the data in [2], [3] and [4]. The determined thresholds are for reference sounds with a degree of modulation m=0.4 in the range of 17-22%, for m=0.5 about 12% and for m=0.6 between 10% and 15%. The influence of additional noise at a SNR between 15 dB and 30 dB on the roughness difference thresholds could not be found.

Investigations with engine sounds show a sensitivity of just-noticeable roughness differences depending on the content of the modulation spectra, influenced by engine roughness as well as psychoacoustic roughness.

References

[1] Terhardt, E.: Über akustische Rauigkeit und Schwankungsstärke, Acustica, Bd. 20, 215-224, 1968.

[2] Fastl, H. and Zwicker, E.: Psychoacoustics, Facts and Models, Springer, Berlin, Heidelberg, New York, 2007.

[3] Vogel, A,: Über den Zusammenhang zwischen Rauigkeit und Modulationsgrad. Acustica 32, 300-306, 1975.

[4] Terhardt, E.: Akustische Kommunikation, Springer, 1997.

[5] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör, doctoral thesis, RWTH Aachen University, 1993.

[6] Hellbrück, J., Ellermeier, W.: Hören, Physiologie und Pathologie, Hogrefe, 2004.

[7] Pentland, A.: Maximum Likelihood Estimation: the Best PEST. Perception & Psychophysics, 28, 377-379, 1980.

[8] Taylor, M. and Creeman C. D.: PEST: Effecient Estimates on probability functions, JASA, 41, 782-787, 1967.

[9] Otto, S.: Vergleichende Simulation adaptiver, psychometrischer Verfahren zur Schätzung von Wahrnehmungsschwellen, Magisterarbeit, 2008.

[10] Leek, M.: Adaptive procedures in psychophysical research, Perception & Psychophysics 2001, 63(8), 1279-1292.

[11] Kaernbach, C.: Adaptive threshold estimation with unforced-choice tasks, Perception & Psychophysics, 68 (8), 1377-1388, 2001.

[12] Sottek, R.: Modelling engine roughness, 2009 SAE Noise & Vibration Conference Proceedings, St. Charles, IL.

[13] Klemenz, M.: Die Geräuschqualität bei der Anfahrt elektrischer Schienenfahrzeuge, doctoral thesis, RWTH Aachen University, 2005.

[14] Attia, F., Okker, A.: Untersuchungen zwischen Rauigkeitsmodellen und Übereinstimmung mit dem subjektiven Gehöreindruck, DAGA 1995, 843-846.