

Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/>



**ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013**

**Psychological and Physiological Acoustics
Session 5aPP: Recent Trends in Psychoacoustics I**

5aPP6. Rating the dieselness of vehicle noise using different psychoacoustic methods

Jakob Putner* and Hugo Fastl

***Corresponding author's address: AG Technische Akustik, MMK, TU München, Arcisstr. 21, Munich, 80333, Bavaria, Germany, putner@tum.de**

Modern diesel engines meet the demand for high power-engines while strict emission regulations have to be fulfilled. Therefore, diesel engines entered vehicle segments where the expectations on the sound quality are exceptionally high. Sound Quality and fuel efficiency are often conflicting goals during the development of a diesel engine. The typical sound character of diesel engines, the so called Dieselness, is an indicator for the overall Sound Quality of the vehicle noise. Hence, it is desirable to rate the Dieselness of engine sounds. Sounds emitted by gasoline- and diesel-powered vehicles in idle condition were rated in psychoacoustic experiments using different methods. First, the method of line length was used as direct scaling procedure to get ratio ratings of the relative Dieselness of the vehicle noises. Second, a direct ranking of the noises has been done with the Random Access method where subjects had to rank the sounds according to their Dieselness. Third, in a paired-comparison test the participants had to judge which of two sounds had more Dieselness, resulting in an indirect scaling. These methods are compared regarding the time the experiments took and the resulting ranking respectively scaling.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

Diesel engines have gained high market shares in almost every category of passenger vehicles over the past years since they meet the demands for fuel efficiency and high-performance. The typical engine sound makes even modern diesel-powered cars distinguishable from their gasoline-powered counterparts. This typical sound character was called Dieselness by Patsouras et al. (2001). The sound of gasoline-powered cars is usually preferred over the sound of diesel-powered cars even if the sound pressure level of the gasoline engine sound is higher, which has been shown by Patsouras et al. (2002).

Several parameters influence the sound emitted by a diesel engine, the settings of the engine control unit can for example alter the sound in a wide range without changing the engine construction. The optimization of fuel efficiency and sound quality are conflicting goals during the vehicle development, which creates the need for efficient rating methods. As mentioned by Fastl et al. (2008), the rating of Dieselness can be a difficult task depending on the rated diesel engine sounds.

In order to find an efficient method to rate Dieselness, three different methods for psychoacoustic experiments were evaluated. The ranking method random access, the direct scaling method of line length and the indirect scaling method paired comparison were chosen for the psychoacoustic experiments. The methods are compared regarding the experiment duration and the resulting ranking respectively scaling.

EXPERIMENTS

Stimuli, Setup and Subjects

Exterior idle noise measurements of vehicles in a semi-anechoic chamber were used for the experiments. A six-cylinder diesel-powered executive car (D) with four different settings of the engine control unit and its gasoline-powered six-cylinder counterpart (G) were measured at microphone positions 2 m in front (F) of the car and in 1 m distance of the car's B-pillar (B). The ten vehicle idle noises have sound pressure levels $L_{AF,max}$ between 50.4 and 64.8 dB(A). For the experiments 1.5 sec of the sounds were chosen and a Gaussian shaping with 50 msec rise and fall time was applied to the sounds.

All experiments were performed in a darkened, sound proof booth. The sounds were presented diotically via electrodynamic headphones (Beyer DT48) with free-field equalization according to Fastl and Zwicker (2007, p. 7). All stimuli were presented with the original sound pressure level.

In all three experiments, the same 12 subjects participated. Their ages ranged from 22 to 32 years (median 24.5 years) and all reported normal hearing.

Procedures and Results

In all three listening tests, the subjects were asked to rate the sound regarding their Dieselness, described as the typical sound character of a diesel engine. Except for a description of the procedure, no further advice was given to prevent biasing the results. The subjects completed the listening tests in a random order, so no specific method would be influenced by learning effects. Between the listening tests the subjects had a longer break of at least 10 min.

Random Access

Random access is a method originally introduced for psychoacoustic experiments by Fastl (2000). The subject's task is to rank the stimuli from low to high Dieselness. For the ranking a graphical user interface is used, where the subjects have random access to all stimuli, i.e. they can listen to the sounds in any sequence and as often as they like. The ranking can be changed iteratively until the subject decides that the current ranking is satisfactory. Each subject rated each stimulus four times. Since the subjects control the test sequence, no trial run is necessary. Typically the first ranking took longer since the subjects had to become acquainted with the stimuli and the procedure. All four rankings with the random access method took between 6 and 17 min (median 11 min).

The results of the random access experiment are shown in Figure 1 as the inter-individual medians and inter-quartile ranges of the intra-individual medians shown as open circles and the inter-individual medians and inter-

quartile ranges of the intra-individual inter-quartile ranges shown as closed circles. Inter-individual interquartile ranges smaller than one rank show very good agreement of the subjects for all ten stimuli. Also the small intra-individual interquartile ranges, typically less than one rank, show very good reproducibility of the rankings for each subject.

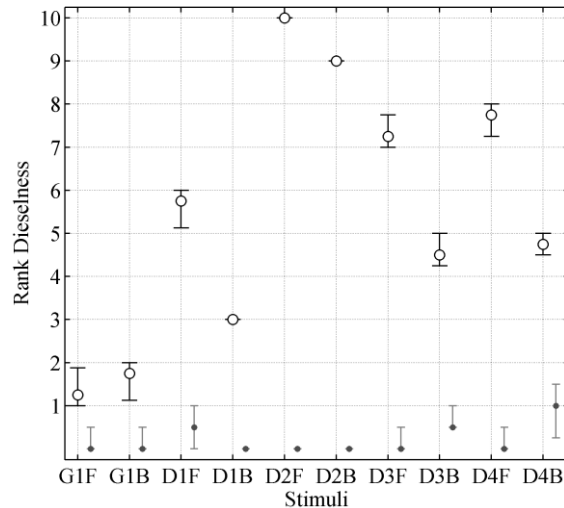


FIGURE 1. Random access Dieselness ranking of exterior idle noises (engine type: Gasoline / Diesel; microphone position: Front / B-pillar). Medians and inter-quartile ranges of intra-individual medians (open circles) and inter-quartile ranges (filled circles).

As expected the two sounds of the gasoline-powered car are rated with the lowest ranks of Dieselness. The microphone in front of the vehicle is generally ranked higher than the microphone near the B-pillar of the same diesel engine configuration. The diesel engine configuration 2 has the highest Dieselness regardless of the microphone position, whereas the other configurations range in the center span.

Line Length Method

As an alternative method for the rating of the Dieselness, the line length method, for example used for psychoacoustic experiments by Fastl et al. (1989), was tested. The subjects had to rate the Dieselness on a horizontal line marked with “very low Dieselness” and “very high Dieselness” at the ends. The line for the rating had a fixed

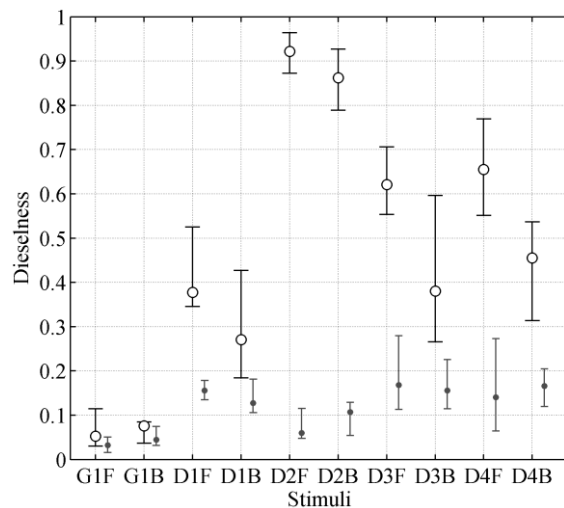


FIGURE 2. Line length rating of the Dieselness of exterior idle noises (engine type: Gasoline / Diesel; microphone position: Front / B-pillar). Medians and inter-quartile ranges of intra-individual medians (open circles) and inter-quartile ranges (filled circles).

length and, aside from the adjective pair at the ends, no further labels. The subjects were asked to rate the Dieselness by pointing at the desired position on the line presented on a touchscreen. The resolution of the touchscreen was high enough to speak of a quasi-continuous scale. The stimulus was presented one time, without the possibility to repeat the stimulus playback, and the subjects had to rate the Dieselness of the stimulus. After a trial period, that consisted of all stimuli in random order, the subjects had to rate each of the randomly presented stimuli twelve times. Despite the large number of repeats, the line length experiment only took between 6 and 11 min (median 7 min).

In Figure 2 the relative line length starting at “very low Dieselness” for each stimulus is shown. Displayed are the inter-individual medians and inter-quartile ranges of the intra-individual medians as open circles and the inter-individual medians and inter-quartile ranges of the intra-individual inter-quartile ranges as closed circles. Rather large partially overlapping inter-individual interquartile ranges show that the subjects as a group did not always rate the stimuli consistently. Intra-individual interquartile ranges up to almost 30 % relative Dieselness show that the subjects had problems to reproduce their own ratings, especially for diesel engine configurations 1, 3 and 4.

However, the gasoline-powered car is rated with the lowest Dieselness and diesel engine configuration 2 with the highest Dieselness. The other engine configurations are rated in between and a tendency that the front microphone is rated higher than the B-pillar microphone position can be seen.

Paired Comparison

In contrast to the direct scaling of the method of line length, the paired comparison method typically uses indirect scaling by representing the choice frequencies with a probabilistic choice model. The subject’s task is to answer which of two signals has more Dieselness i.e. sounds more like a diesel-powered car. The two signals were reproduced with a 125 msec pause in between and the subjects had to reply by pressing the left arrow key if the first signal was more likely to be a diesel-powered car or by pressing the right arrow key if it was the other way round. There was no possibility to repeat the signal pair. For each of the 45 pairs of the 10 stimuli it was assured that each stimulus appeared equally often on the first and second position following a balanced test design. A short trial of five pairs preceded the test. The paired comparison experiment took between 15 and 19 min (median 16 min).

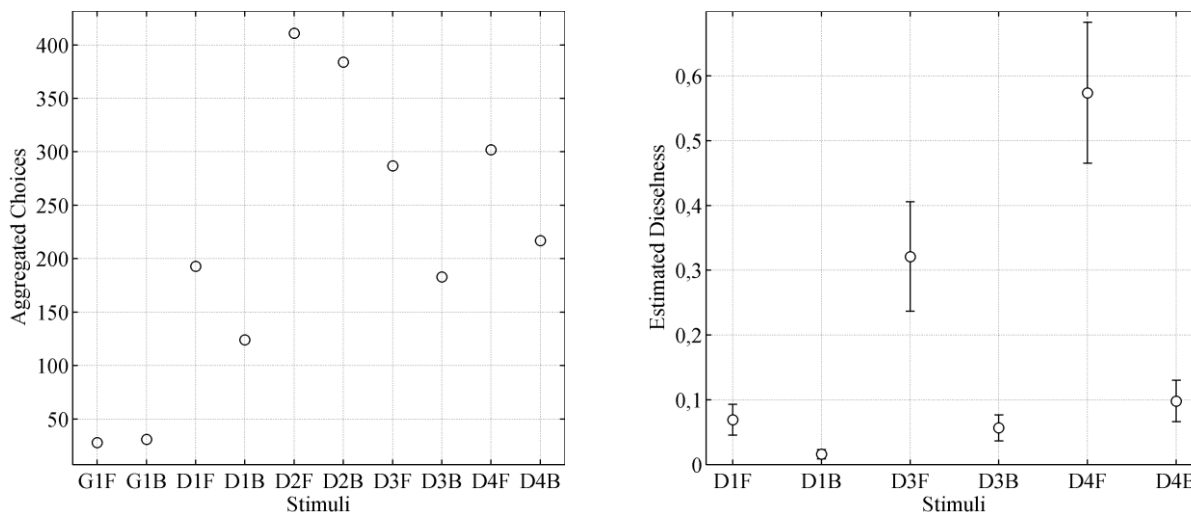


FIGURE 3. Results of the paired comparison experiment (engine type: Gasoline / Diesel; microphone position: Front / B-pillar): The values of the aggregated choices (left) and parameter estimates of the Bradley-Terry-Luce model with approximate 95 % confidence intervals ($\chi^2(10) = 4.97$, $p = 0.107$) (right).

A common model for paired comparison data is the Bradley-Terry-Luce (BTL) model by Bradley and Terry (1952) and Luce (1959). However, no BTL model could be estimated that fits the data of all stimuli good enough. Since some of the stimuli have similarities, i.e. diesel or gasoline engines, engine control unit parameters or microphone positions, several Preetree models according to Wickelmaier and Schmid (2004) have been evaluated. The Preetree models group the similarities, like engine type, on their branches by adding additional group parameters. However, this approach did not lead to a model with a good fit. Since stimuli that are never or always preferred may lead to problems with the BTL model estimation, the data of some stimuli was excluded from further

analysis. From the aggregated choices in Figure 3 (left), stimuli that were almost always rejected or preferred can be identified. For the aggregated choices all decisions for a stimulus are added. The gasoline-powered car was generally never rated higher than the diesel-powered cars, indicating the lowest Dieselness of all sounds. Diesel engine 2 however, was almost always preferred over the other stimuli, indicating a very high Dieselness. If the gasoline-powered car and diesel engine 2 are removed from the analysis, a BTL model with a good fit can be estimated. The estimated Dieselness and the 95 % confidence intervals are shown in Figure 3 (right). Besides the sounds of the microphone position in front of the car for diesel engine configuration 3 and 4, which show higher estimated Dieselness, all other sounds show quite similar values.

DISCUSSION

All three methods provide reasonable results for the Dieselness of the car exterior idle noise. Since the BTL model can only estimate the Dieselness of six sounds on a ratio scale, the aggregated choices are used for a comparison of the methods with all stimuli. In Figure 4 the very good accordance between the values can be seen. One has to keep in mind the different scales of measure, the results of the random access method are on an ordinal scale and the aggregated choices from the paired comparison test also. The method of line length on the other hand directly provides ratio scaled values. However, the inter-individual inter-quartile ranges of the method of line length seem much larger than the corresponding inter-individual inter-quartile ranges for the random access method, indicating a higher uncertainty between the subjects. Also the inter-individual medians and inter-quartile ranges of the intra-individual inter-quartile ranges, shown in Figure 1 and 2, indicate a better reproducibility of the responses of one subject.

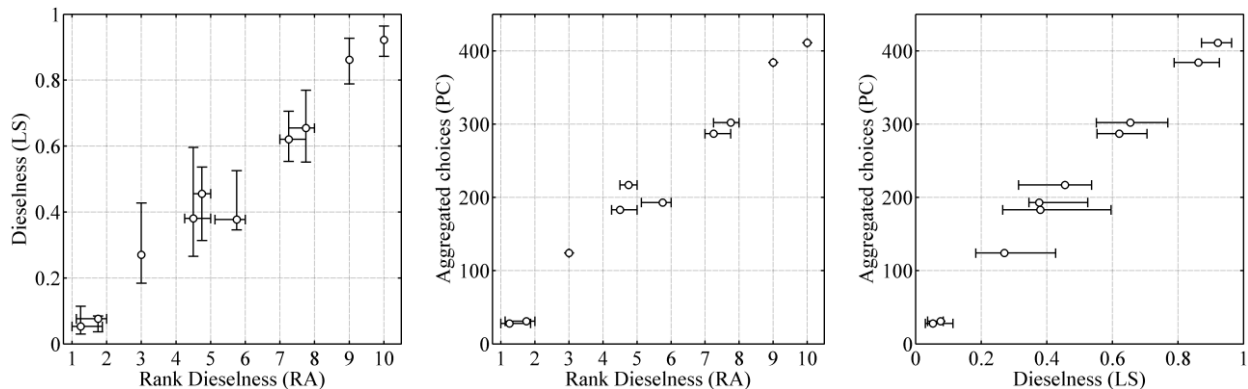


FIGURE 4. Comparison of the results of the different psychoacoustic methods for all sounds: Random access (RA) and Line Scaling (LS) ($\tau = 0.91$) (left); Random access (RA) and Aggregated Choices (PC) ($\tau = 0.96$) (middle); Line Scaling (LS) and Aggregated Choices (PC) ($\tau = 0.99$) (right).

For the paired comparison method, ratio scale values for six stimuli can be estimated using the BTL model. These values are compared to the corresponding six values of the other methods in Figure 5. It has already shown during the discussion of the results, that these six stimuli have smaller differences in the perceived Dieselness than the excluded stimuli. The switched ranks between the random access and the line length results are more evident in Figure 5 (left) than in Figure 4 (left). Otherwise the rank correlation between the results is still very good. The confidence intervals of the BTL model are rather small, but the comparison to the other results shows differences. While the comparison to the aggregated choices in Figure 4 showed very good accordance, the data points approximate the bisecting line, the values in Figure 5 show larger differences. For the comparison of the ratio scaled values of the method of line length and the BTL model a better accordance would have been expected. This possibly hints that more subjects need to participate in the paired comparison test for a better fit of the BTL model.

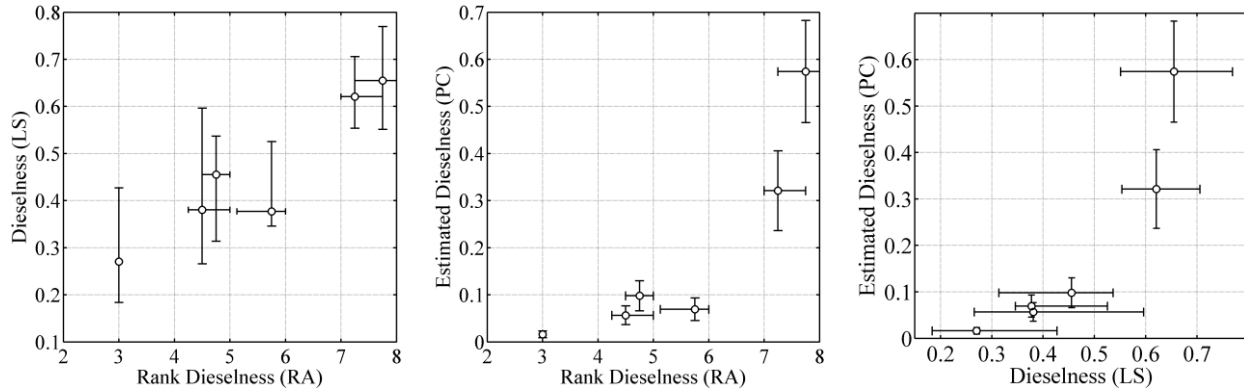


FIGURE 5. Comparison of the results of the different psychoacoustic methods for the six selected sounds: Random access (RA) and Line Scaling (LS) ($\tau = 0.73$)(left); Random access (RA) and Estimated Dieselness (PC) ($\tau = 0.87$)(middle); Line Scaling (LS) and Estimated Dieselness (PC) ($\tau = 0.87$) (right).

The durations of the different experiments are summarized in Table 1. A comparison of the average durations shows that the method of line length is the fastest and that a paired comparison test takes most time. But the small variability of the durations of the paired comparison test shows that the duration can be estimated a priori using the number of pairs and their duration. The durations of the random access test show a higher variability, but a connection between short durations and poor reproducibility of the subject’s rankings could not be found.

TABLE 1. Durations of the different psychoacoustic experiments in minutes.

Method	Minimum	Median	Maximum
Random access	6	11	17
Line length	6	7	11
Paired comparison	15	16	19

CONCLUSION

Three methods for the rating of the typical sound of a diesel engine, the so called Dieselness, have been presented and evaluated. The resulting Dieselness ratings of all methods were plausible and quite similar, yet the scale of measure, the reproducibility of the ratings of a single subject, the agreement of the subjects and the test time differed.

The random access method, where subjects perform a direct ranking of the stimuli, provided repeatable results and very good agreement of the subjects in relatively short time. However it has to be noted that the results are on an ordinal scale, but the interquartile ranges can give hints about relations between the stimuli. Whenever the ratio scale is not necessary, random access seems to be the method of choice for psychoacoustic experiments evaluating Dieselness.

Ratio scaled values are provided by the method of line length, for which the subjects rate the Dieselness directly on a line, in very short test durations. The downsides are a decreased reproducibility of the rating of a single subject and increased disagreement between the test subjects leading to an uncertainty in the results.

Paired comparison tests have shown to be very time consuming since a single experiment takes very long and a larger group of subjects is needed for reliable ratio scale estimates from a Bradley-Terry-Luce (BTL) model. Also the selected stimuli have huge influence on the fit of the BTL model and likely the number of required subjects for the experiment. If ratio scale values are needed, the paired comparison method is a useful tool, but at extensive costs, since the experiments are time consuming and a large group of subjects has to participate. It also seems advisable to perform faster experiments prior to the paired comparison test in order to exclude extreme values from the experiment.

ACKNOWLEDGMENTS

This research was supported by the Bavarian Research Foundation as part of the FORLärm research cooperation for noise reduction in technical equipment.

REFERENCES

- Bradley, R. A. and M. E. Terry (1952). "Rank analysis of incomplete block designs: I. The method of paired comparisons.", *Biometrika*, **39**, 324–345.
- Fastl H., E. Zwicker, S. Kuwano, S. Namba (1989). "Beschreibung von Lärmimmissionen anhand der Lautheit", In: *Fortschritte der Akustik, DAGA'89*, 751–754 (DPG, Bad Honnef).
- Fastl, H. (2000). "Sound Quality of Electric Razors - Effects of Loudness", In: *Proc. inter-noise'2000*, Nice, France.
- Fastl, H., and E. Zwicker (2007). *Psychoacoustics. Facts and Models*, 3rd ed. (Springer, Berlin).
- Fastl, H., B. Priewasser, M. Fruhmann and H. Finsterhölzl (2008). "Rating the Dieselness of engine-sounds", In: *Proc. Acoustics 08*, Paris, France, 1021–1024.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*, (Wiley, New York).
- Patsouras, C., H. Fastl, D. Patsouras and K. Pfaffelhuber (2002). "Psychoacoustic sensation magnitudes and sound quality ratings of upper middle class cars' idling noise", In: *Proc. 17. ICA Rome*, Rome, Italy.
- Patsouras, C., H. Fastl, D. Patsouras and K. Pfaffelhuber (2001). "How far the sound quality of a diesel powered car away from that of a gasoline powered one?", In: *Proc. Forum Acusticum Sevilla 2002*, Sevilla, Spain.
- Wickelmaier, F. and Schmid, C. (2004). "A Matlab function to estimate choice model parameters from paired-comparison data", *Behavior Research Methods, Instruments, & Computers*, **36**, 29–40.