

Analytical Analysis of Disturbed Radio Broadcast

Jan Reimes¹, Marc Lepage¹, Frank Kettler¹
Jörg Zerlik², Frank Homann², Christoph Montag²

¹HEAD acoustics GmbH, 52134 Herzogenrath, Germany

²Robert Bosch Car Multimedia GmbH, 31139 Hildesheim, Germany

jan.reimes@head-acoustics.de

Abstract

Analogue radio broadcast signals are often disturbed in various ways as experienced in driving situations. These disturbances cover typical, short term "pops", longer noise bursts, short term "mutes", insertion of so-called "HighCut", i.e. the attenuation of higher frequencies to mask disturbances or noticeable stereo/mono switching. These disturbances can be separated in two parts: they are caused either by bad reception conditions (loss of transmission energy, interferences, multiple transmission paths etc.) or by implemented masking techniques in FM receivers in order to improve audio quality.

Index Terms: radio broadcast, quality of music, auditory tests, analytical model

1. Introduction

It is the aim to derive an analytical method to examine the occurrence of typical artifacts of disturbed radio broadcast transmission and their effect on the quality of the signal perceived by human listeners. A reproducible method for quality assessment is therefore highly desired. As this method shall model the human perception of the disturbances, auditory tests needed to be conducted first. First results have been presented [1] and were used to derive a data base to develop such an analytical model. The model in its current form matches these auditory data as shown in [2]. During further development more auditory tests were conducted using real life recordings, i.e. receiver recordings made during test drives on public roads under most realistic conditions. This data basis is now used for further tuning and adaption of the model. Results from the auditory and analytical tests are discussed in this contribution.

2. Auditory and Analytical Testing

2.1. General Considerations

Various aspects needed to be considered in the design phase of the auditory tests, such as

- the choice of appropriate audio samples (temporal structure, dynamic behavior)
- the generation of reproducible but realistic disturbances, both in isolated form and as combined disturbances (multiple disturbances)
- the consideration of the correct acoustical environment in the listening situation
- generation of a realistic cognitive load in the listening situation for subjects during the tests.

The principle of the test design is shown in figure 1. The selected audio samples were processed and realistically disturbed.

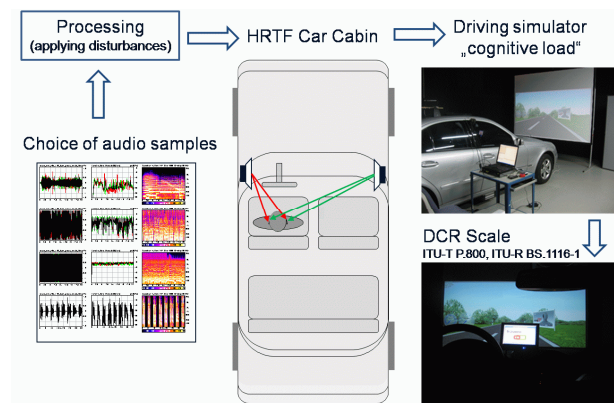


Figure 1: Principle test setup for auditory tests.

The head-related transfer function (HRTF) filtering guarantees the consideration of the acoustic environment, the car cabin. The convoluted samples are then presented to test persons in an interactive driving simulator to simulate a realistic listening environment and an appropriate cognitive load for the subjects during the tests. The quality rating is given on a Degradation Category Rating scale (DCR, [3]) via touchscreen in the car cabin. A brief overview about the auditory test is given in the next sections. More details about this evaluation can be found in [1].

2.2. Choice of Test Samples

The tests were conducted using three different music samples and two speech sequences. The sample duration was chosen to 20 seconds each. The music samples were selected according to the following, signal dependant criteria:

- Rock music: Dense arrangement, rhythmic and loud, using drums and distorted electric guitars. This music style is supposed to mask many artefacts, especially those with highly varying, temporal structure.
- Pop music: less dense and less loud than rock music, also rhythmic, using intense stereo effects, with vocals.
- Classic music: Quiet passages, sustained piano tones, orchestra sound.
- Speech: approved English samples according to ITU-T P.501 [4] were selected.

All music samples featured full audio bandwidth (20 Hz to 20 kHz). The English speech samples were of up to 16 kHz bandwidth.

2.3. Single Disturbances and their Generation

The audio files were exposed to individual disturbances to be judged in the listening test. Five different kinds of typical disturbances were considered, generated as follows:

- "Pops": Pop noises represent short term noise bursts mainly caused by multiple paths interferences. For the test, the signals were transmitted via an RF simulation with multiple transmission paths. The output of an FM receiver was recorded. The receiver was configured to suppress all other kind of signal processing to improve quality.
- "Noise": In contrast to the "Pops", "Noise" is characterized by a longer duration (> 0.5 seconds) of the noise disturbance. The noise signal was generated by dynamically fading an unmodulated RF carrier. Later the resulting noise was mixed with the unaffected audio signal.
- "Mutes": The signal is partially attenuated or totally muted. The disturbance was simulated using an audio editor.
- "HighCut": The signal is temporarily lowpass filtered. The examples were also generated with an audio editor.
- "Mono/Stereo": The stereo width of the signal is reduced temporarily, again by using an audio editor tool.

All impairments were generated in several grades, simulating light, medium and strong disturbances. The test samples were completed by a set of multiple disturbed files, generated by applying a combination of the aforementioned methods and also by recordings done in the field.

2.4. HRTF Filtering and Driving Simulation

The influence of a real car cabin was captured by measuring the HRTF of the vehicle including its sound system. For this purpose, an artificial head measurement system was placed on the driver's seat of the car to be evaluated. A maximum length sequence (MLS) was fed into the audio system of the car and recorded by the artificial head's left and right ears. From these recordings, the following HRTFs were calculated: left speakers to left ear, left speakers to right ear, right speakers to left ear and right speakers to right ear. All test signals were convoluted with these impulse responses to imprint the characteristics of the acoustic car environment.

To gain a realistic cognitive load for the test person during the evaluation, an acoustical subsystem of a driving simulator incl. visual feedback system was used. The driving simulator boasts a virtual dynamic driving model coupled with a video system that simulates the movement of the car on a straight street. The sound system of the simulator was disabled. The test persons wore diffuse field equalized open headphones in order to best reproduce the sounds impression after the HRTF filtering.

2.5. Test Signal Levels

According to the ITU-R Recommendation BS.1116 [3], the recommended overall test signal level for audio quality tests should be set to 82 dB(A). However, all participants involved in the project found that to be much too loud for the situation to be tested. A pre-test described in [1] leads to 65 dB(A) as an appropriate playback level, which was also used for the presentation in the listening test.

2.6. Test Realization

18 subjects participated in the listening tests. Test persons were aged 23 to 52 years, 15 male and 3 female, all with normal hearing as proved by audiometry prior to the tests. Six of the subjects were "expert listeners" involved in the project; the other twelve were naïve test persons. The test persons were given the following instructions:

"Imagine, you are driving home after work and listening to the radio. It provides music that you like or news that you're interested in. Both can either be not disturbed at all or partially disturbed. Please rate the impairments according to the following scale:

*Impairments are... Imperceptible
 Perceptible, but not annoying
 Slightly annoying
 Annoying
 Very annoying"*

This scale is according to the ITU-R BS.1116 [3] and ITU-T P.800 [5] with the score 5 being the best rating ("Imperceptible"). The lowest scale ("very annoying", score 1) should describe a situation, where the subject would definitely choose a different broadcast station.

2.7. Test Results

110 conditions including undisturbed references were judged by the subjects, leading to a total number of around 2000 single judgements. An overall view on the average results versus all conditions is presented in figure 2a and 2b.

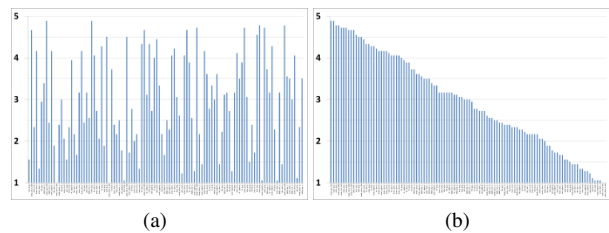


Figure 2: LOT results in order of (a) appearance during the test and (b) sorted in descending order.

Figure 2a shows the results in order of presentation of the listing examples in the auditory test. The presentation order of the listening examples was equally randomized. In contrast, figure 2b sorts the results by descending MOS scores. The scores are equally distributed over the entire MOS range, which is very helpful if the data serve as basis for the development of an analytical model. The whole DCR scale was covered during the test. All averaged scores feature a confidence interval of ≤ 0.5 MOS (not shown in figure 2a and 2b for reasons of clearness).

Almost all test results show decreasing scores for increasing disturbances, as expected. For the mono/stereo switching conditions, the range of auditory scores are quite narrow (around 4.0 MOS) or is influenced by level loss. This indicates that this type of disturbance has only a minor perceptual influence. Thus, this disturbance type is excluded for the usage in an objective predictor in a first step. The detailed results (grouped per disturbance and per test sample) and further descriptions are published in [1].

3. Objective Analysis

The subjective testing described in section 2 is used as a basis for the development of an objective, hearing-adequate model which is able to predict the perceived quality of a given recording on a MOS scale. Such an analytical model shall be capable to analyse the implemented signal processing in receivers, consider human perception of typical disturbances, provide results with high correlation and consequently needs to be based on auditory tests in realistic environments (incl. the acoustics of the vehicle cabins). The analytical approach presented is reference-based, means that a best-case recording is available and can be used for comparison.

3.1. Pre-Processing

Several steps of pre-processing have to be considered before applying the actual analysis to the recordings:

- The degraded and the reference signals need to be time aligned, thus an overall delay compensation needs to be applied.
- HRTF filtering is applied in order to simulate the acoustic environment of a target vehicle cabin.
- The levels of the reference and the degraded signal need to be comparable. Therefore, the loudest channel of both signals is adjusted to a reference level of 65 dB(A) (see section 2.5). The level difference between the left and the right channels is kept unchanged.
- The input files for the analyses usually result from recordings with a length of several minutes. These recordings need to be segmented in order to be applicable for auditory testing. Segments of 20 s length have been chosen, which is identical to the listening test procedure.
- It has to be taken into account that receivers under test can provide vehicle specific equalizations which cannot be deactivated for testing. To compensate this influence, a long-term transfer function (over the full, unsegmented measurement) between the reference and degraded signal is determined. This filter is applied to the *reference file* in order to make the reference more similar compared to the degraded signal.

3.2. Detection of Single Disturbances

The basis for the overall quality estimation of the objective model is the capability to detect the typical disturbances described in 2.3.

3.2.1. Mute

A "mute" (or attenuation) of the audio signal is introduced by the receiver in case of RF signal loss in order to avoid noise that would occur otherwise. The resulting gaps in the time signal are exemplarily shown in the upper diagram in figure 3.

A relatively easy way to analytically detect these "mutes" is a channel wise delta-level vs. time analysis between degraded and reference signal. The result of this analysis is shown in the lower diagram in figure 3. The four "mutes" with different characteristics are clearly detected.

From this analysis single value descriptors can be derived and mapped against the auditory test data. This is done with a linear, multiple regression of parameters, which were identified

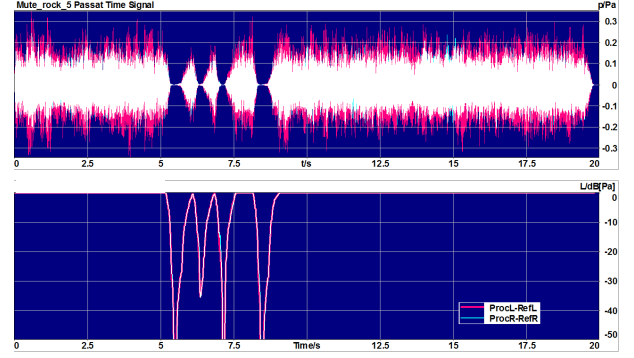


Figure 3: Principle of mute analysis: delta-level vs. Time

to best describe the relation between the MOS score and the extracted parameters. For the 24 different "mute" conditions covered by the listening test this leads to a rather high correlation result of 95.3%.

In a next step, the analytical quality scores are calculated for the remaining 87 non-"mute" conditions using the trained method as derived based on the "mute" conditions. The results displayed in the left diagram in figure 4a are achieved.

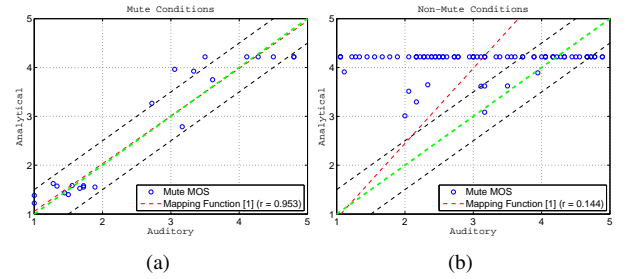


Figure 4: Correlation of "mute" training data (a) and application on non-"mute" files (b)

It can be seen that the "mute" detection is relatively robust against non-"mute" disturbances (see figure 4b). Most of the conditions are rated correctly with the maximum achievable score of 4.2 MOS (highest possible score in the training). Only a small subset of non-"mute" conditions is rated lower than this maximum. This can be explained by other disturbances introducing strong attenuations in the considered files (e.g. strong "HighCuts" or "Mono/Stereo switches") resulting in larger overall attenuations which are detected by the "mute" analysis.

3.2.2. HighCut

Low-pass filtering is inserted by the device under test (DUT) to reduce eventual noise / pops artifacts. This masking technique also influences the frequency shape of the audio signal and thus the original sound. The base analysis for the identification algorithm is a smoothed 12th octave vs. time spectrogram, which is applied on both channels of the reference and degraded signal. A delta-spectrum (degraded – reference) is created for left and right channel independently; an example is given on the left side of figure 5. In order to compare "HighCuts" (temporal insertion of low pass), the upper frequency band level vs. time (12kHz - 16kHz) is extracted from this delta-spectrum to detect high cuts. To distinguish between "HighCuts" and "mutes", addition-

ally the lower frequency band level vs. time (100Hz - 1kHz) is extracted. The negative peaks of the high-band level vs. time curve (see right side of 5) are only detected as "HighCuts" when the low-band delta-level vs. time remains unchanged at 0 dB.

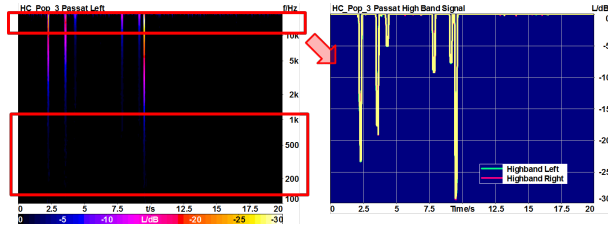


Figure 5: Generation of low- and high-band delta-level vs. time

Similar as for the "mute"-detection, numeric values can be extracted from the curves to quantify the impact of the "HighCuts". Figure 6a shows the prediction results in the same way as in the previous section. Again, the so-called "orthogonality" against the other disturbances is given, as shown in figure 6b. Nevertheless, as already noticed in [1], the auditory impact of the "HighCuts" disturbances is rather low. Most of the conditions obtained MOS values > 3.0 MOS.

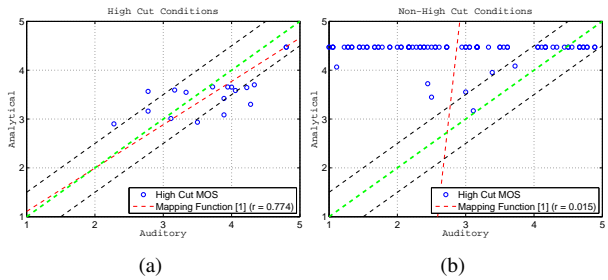


Figure 6: Correlation of "high cut" training data (a) and application on non-"high cut" files (b)

3.2.3. Noise

Low SNR conditions / bad demodulation can lead to additive noise. Usually this artifact is masked by muting the output or inserting temporal low-pass, but in some cases this disturbance can occur.

To detect and quantify the noise components within the signal, again a delta-analysis is chosen. The noise estimation is carried out on the degraded spectrogram with the so-called *Minimum Statistics* approach [6]. In conjunction with the known reference and its noise estimation determined by the same algorithm, the delta between both noise-spectrograms is calculated (see figure 7). Based on this delta-spectrogram, the average over frequencies is applied which results in a curve "delta-noisiness vs. time". With this analysis, only noise which was not already present in the reference signal is detected.

Similar as for previous analyses, numeric values can be extracted from the curves to quantify the impact of the "noise" disturbance. Figure 8a shows the prediction results in the same way as for the other disturbances. Again, the so-called "orthogonality" against the other disturbances is given, as shown in figure 8b.

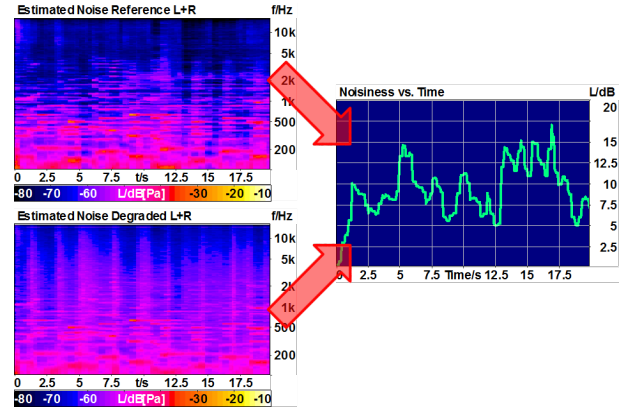


Figure 7: Principle of noise analysis: Calculation of "Delta-Noisiness vs. time"

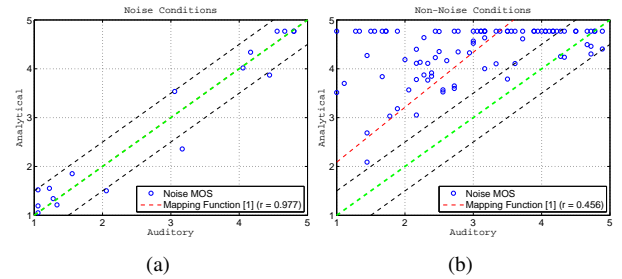


Figure 8: Correlation of "noise" training data (a) and application on non-"noise" files (b)

3.2.4. Pops

The dominating type of disturbances in radio broadcast transmission is the so-called "pops". These short noise bursts result from low short-time SNR conditions or bad demodulations at the radio broadcast receiver. In order to detect those short noise burst occurrences the Relative Approach analysis is used [7]. The Relative Approach was originally designed for detecting spectral or temporal patterns ("noticeable features") based on the capability of human hearing to adapt itself to an expected sound event. The 3D-Relative Approach representation of the undisturbed reference signal is shown in the upper left diagram in figure 9. The same analysis applied to a signal degraded by "pops" occurrences is shown in the lower left hand diagram.

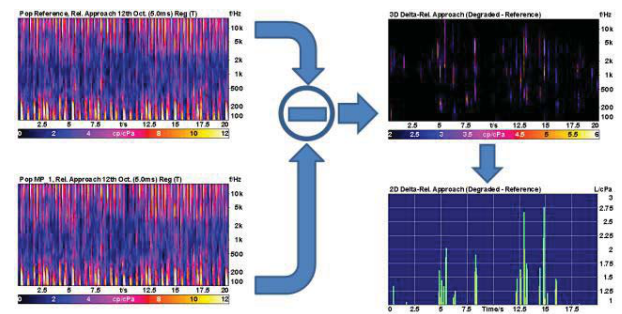


Figure 9: Example of "pops" analysis using the Δ -Relative Approach

At first sight, no big differences can be detected comparing both diagrams. This could be expected since the Relative Approach is designed to be applied to stationary signals in order to detect unexpected spectral or temporal patterns. However, combining both single Relative Approach results using a delta-calculation (subtraction), the result in the upper right hand diagram is provided. In this 3D representation vs. time and vs. frequency the additive disturbances ("pops") are clearly highlighted. Furthermore, if the Δ -Relative Approach values are averaged over all frequencies this results in the two-dimensional representation in the lower right hand diagram. In this representation the temporal distribution of the "pops" occurrences is clearly displayed.

A single value quality score can be extracted from these representations using statistical calculation. Analog to the preceding analyses these single values can be mapped to auditory test results through a training phase. The correlation results are shown in figure 10a, together with the correlation to all other non-"pops" conditions in figure 10b. Here the orthogonality is not as obvious as for the other analyses, which is caused by the other disturbances. In consequence, the analysis of "pops" disturbances in current state is not yet a fully reliable measure. Nevertheless, it still can be used for the overall quality prediction (see next section 3.3).

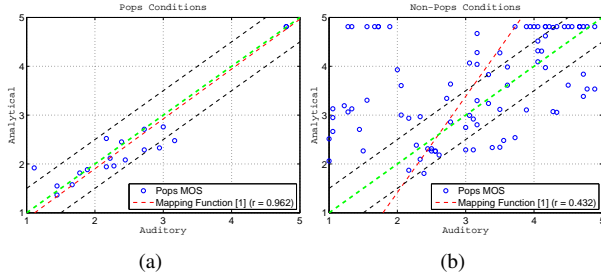


Figure 10: Correlation of "pops" training data (a) and application on non-"pops" files (b)

3.3. Composition of Overall Quality

As shown in the preceding sections for the examples of "mute" and "pops" disturbances, quality parameters can be extracted from different analytical approaches in order to reproduce auditory test results. This can be done i.e. for "mutes", "HighCuts", "noise", "pops" degradation and stereo/mono switching. The idea of separate orthogonal analyses for all of these degradation types is to provide a quality score representing only the degradation introduced by a specific type of disturbance. Additionally, the single disturbance MOS can be combined to an overall quality score e.g. using a neural network according to [8] trained on the totality of the auditory data as indicated in figure 11. Since the stereo/mono switching was found not to lead to strong audible degradations during the auditory tests (see also section 2.7), a single MOS for this type of disturbance is not yet provided and does not contribute to the calculation of the overall MOS.

Correlation results for the analytically derived overall MOS compared to auditory data from the third party listening test are shown in figure 12a and b.

If the complete data set used in the listening test is considered, an overall result as represented in figure 12a can be achieved.

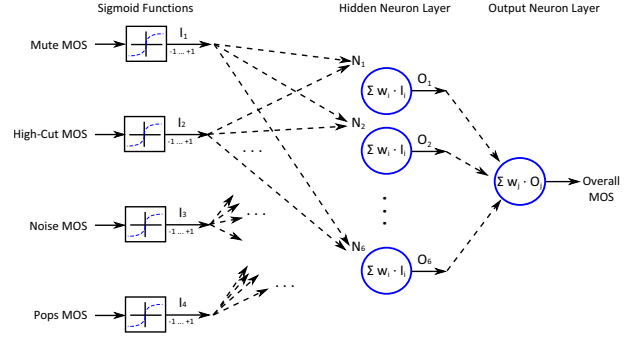


Figure 11: Structure of neural network used for result combination

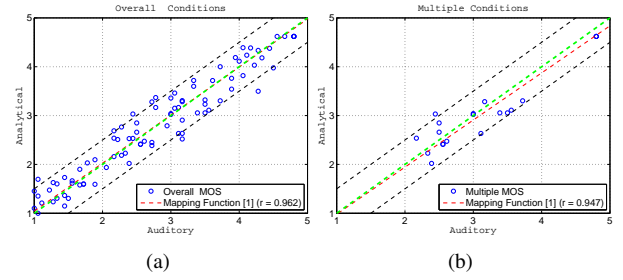


Figure 12: Overall analytical MOS vs. auditory MOS for entire training database (a) and subset of multiple disturbed files (b)

In figure 12b the correlation results for a subset of multiple disturbed conditions (more than one type of degradations) are represented. The high correlation indicates a high potential for the selected analytical approach to reproduce auditory test results.

4. Validation with Real-Life Recordings

In order to validate the model with unknown data, a second listening only test was conducted. The test environment and procedure were identical with the setup described in section 2. The new test contained 50% of the described audio material and 50% real-life recordings (and thus real-life broadcast station audio data). 50% of this dataset of each category is used for validation, the other half should remain for further development.

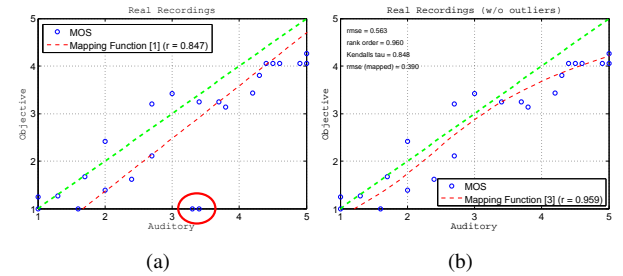


Figure 13: Correlation results of test 2: (a) full subset and (b) outlier removed

Figure 13 shows the raw output scores of the objective model. On the left side (13a), all 25 real-life conditions are presented. The correlation coefficient of 84.7% seems not to

perform adequately. But due to the small amount of conditions, the results dramatically increase when only removing the two largest outliers (see red circle in figure 13a). These results are shown in figure 13b and achieve a correlation of 95.6% after 3rd order mapping.

5. Conclusions

An auditory listening test was conducted to quantify the annoyance of typical artifacts present in FM reception of disturbed radio broadcast. Around 2000 single judgments were evaluated covering 110 conditions. The results are robust concerning their low confidence intervals and cover the complete quality range on the DCR scale.

The wide range of different audio sample characteristics and typical disturbances, the realistic test infrastructure in a driving simulator and the consideration of the acoustical properties of the car cabin during the tests guarantee realistic and reliable results.

Furthermore, the assessment of isolated, individual disturbances as well as the rating of combinations of multiple disturbances provided a very good data basis for the development of the analytical model to automatically rate the audio quality of FM receivers in bad reception conditions. Nevertheless, the detection of "pops" disturbance has to be optimized in further work in order to support better orthogonality of the single analyses.

The proposed model was tested against unknown and real-life data. Even though some outliers still have to be analyzed, the performance is already promising in order to use the new approach for further quality assessments of broadcast receivers.

6. References

- [1] U. Müsch, M. Lepage, F. Kettler, J. Zerlik, F. Homann, and C. Montag, "Test environment for realistic listening evaluation of disturbed radio broadcast," in *DAGA, Merano, Italy*, 2013.
- [2] M. Lepage, J. Reimes, F. Kettler, U. Müsch, J. Zerlik, F. Homann, and C. Montag, "Analytical hearing-adequate assessment of disturbed radio broadcast," in *DAGA, Merano, Italy*, 2013.
- [3] *BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, ITU-R Recommendation, 10/1997.
- [4] *P.501: Test signals for use in telephony*, ITU-T Recommendation, 09/2009.
- [5] *P.800.1: Mean Opinion Score (MOS) terminology*, ITU-T Recommendation, 07/2006.
- [6] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," in *IEEE Trans. Speech and Audio Processing*, 2001, pp. 504–512.
- [7] K. Genuit, "Objective evaluation of sound quality based on a relative approach," in *Inter-Noise, Liverpool, England*, 1996.
- [8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.