An Approach for Instrumental Quality Evaluation of Car Audio Systems

Magnus Schäfer

HEAD acoustics GmbH, 52134 Herzogenrath, Deutschland, E-Mail: telecom@head-acoustics.de

Abstract

The expected sound quality of car audio systems has continually increased in recent years. Systems with particularly high quality for music playback can be found in the luxury class where the audio system is also used as a marketing tool and a unique selling point. No instrumental quality measures which incorporate a perceptually motivated model for the analysis of the complex sound field in the car are available so far for this application scenario.

An important aspect in the assessment of perceived audio quality in such a multi-channel scenario is the evaluation of the spatial properties of the sound field. A human listener is easily capable of judging audio systems with respect to their spatial fidelity even if all other characteristics are similar.

An approach for the quantification of the spatial properties of the sound field is presented in this contribution which is based on a binaural hearing model. Several metrics are derived from a coincidence-based model and evaluated with respect to their perceptual relevance. The evaluation is based on a recently conducted listening test which consists of music recordings that were made in different cars. A possibility for instrumental assessment is investigated by combining the devised metrics in a regression approach. It is shown that the spatial properties alone are not sufficient for assessing the perceived audio quality.

Introduction

The ground truth for evaluating the performance of an audio system (or any audio signal processing and reproduction system for that matter) is to conduct a listening test which should be sufficiently large to get meaningful statistical results. However, this is fairly time-consuming and cumbersome. Thus, appropriate techniques for instrumental quality evalution are advantageous both for comparing or improving existing systems and for developing new systems.

It was already investigated in [1, 2] how, in certain cases, including a model for spatial perception can improve the performance of instrumental measures, e.g., [3]. The signals that were used in [1], however, only consisted of artificial signal modifications and codec distortions – no acoustic recordings were included. Additionally, it was hitherto not evaluated if the spatial properties alone can be used for predicting the perceived audio quality.

Both areas are addressed in this contribution. Firstly, the presented investigations use real-world recordings in complex acoustic environments and a new auditory dataset. Secondly, the devised instrumental assessment exclusively utilizes the output of a binaural hearing model for quantifying the performance of the audio system. Note that the binaural hearing model is implicitly susceptible to certain non-spatial aspects of the transfer function of the audio system, e.g., changes in the spectral shape. However, the non-spatial characteristics are not explicitly considered here.

Listening Test

An auditory evaluation was recently conducted that aims at comparing the overall performance of different car audio systems in a fair manner. A short description of the listening test is given here, for more details on the test itself and the evaluation of the test results, see [4].

In total, the listening test contained 161 items representing different audio samples, different acoustic environments, different recording conditions and some artificial signal modifications, e.g., bandpass filtering to define anchor conditions. 45 test subjects participated and each listened to all 161 stimuli. Two different test environments (driving simulator and listening laboratory) were used to evaluate the impact of the test environment on the auditory results. No relevant differences were found, consequently, the results from both test environments were joined into one large dataset for the investigation in this contribution.

Binaural Hearing Model

The instrumental assessment in this investigation is based on a binaural hearing model. The model is based on the work of Lindemann in [5, 6] with extensions that were presented in [7].

A detailed description of the underlying structure of the model is beyond the scope of this contribution, only a brief overview of the core elements is given here. The inner ear is modeled by a Gammatone filterbank [8] and a model for the haircell response that was devised by Lindemann. The resulting neural signals are fed into chains of delay elements that are used to calculate the crosscorrelation between the signals in each of the 36 frequency channels for each lateralization. This crosscorrelation is inhibited, i.e., as the signals are propagating through the delay chains, they are continuously attenuated. This attenuation depends on the amplitude of the oncoming signal and the current values of the inhibited crosscorrelation at neighbouring lateralizations.

The output of the core hearing model is the inhibited crosscorrelogram in the 36 frequency bands. Subsequently, the correlograms are weighted both on the lateralization axis as well as on the frequency axis with the weighting functions given in [2] and then averaged across frequency to get the final correlogram $\Psi(k, m)$ (with k denoting the discrete time and m denoting the lateralization). Example correlograms for one audio signal in different environments are given in Figures 1 to 3.

An audio signal of approximately 10 s was analyzed for lateralizations from −1 ms to 1 ms which roughly corresponds to the lateralization range for 180°.

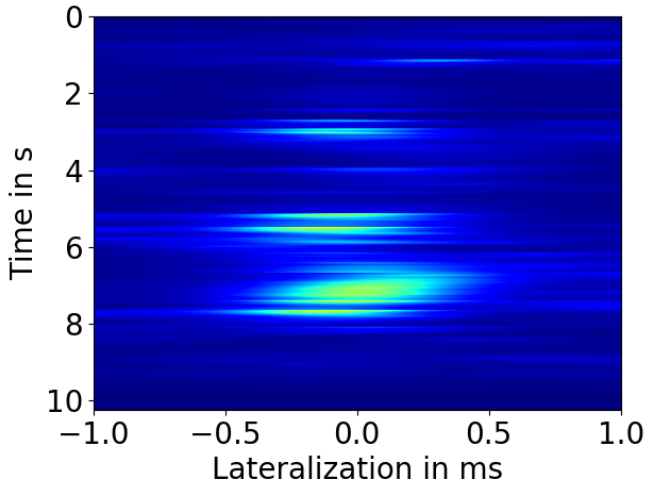

Figure 1: Correlogram for reference signal

The correlogram for the reference sample in Figure 1 gives an indication of both the temporal and the spatial structure of the signal: Most of the activity can be observed at 3 s and then from 5 s to 8 s. Here, the spatial distribution has its centroid located close to a lateralization of 0 ms, i.e., in the middle.

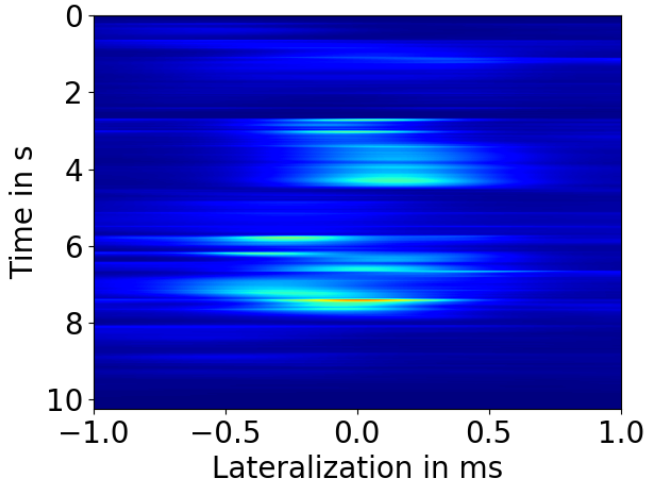

Figure 2: Correlogram for signal with good auditory result

The correlogram for an audio sample that was assessed with a high rating in the listening test is given in Figure 2. While there are some changes in the temporal structure, in particular around 3 s, the overall correlogram is fairly similar to the reference case and the spatial situation did not change dramatically.

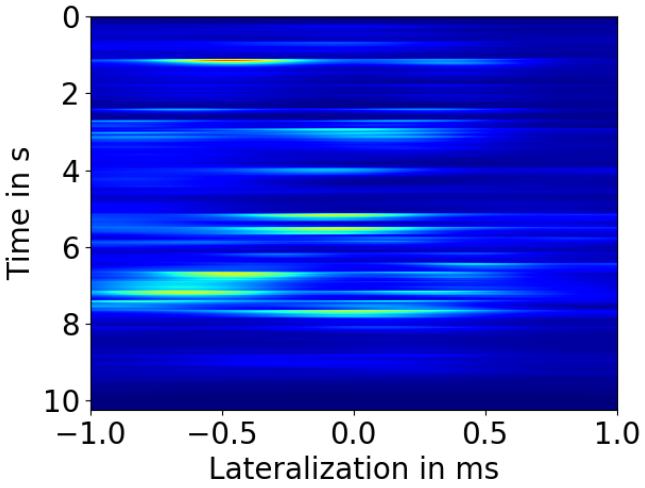

Figure 3: Correlogram for signal with poor auditory result

Figure 3 shows the correlogram for an audio sample that was rated poorly in the listening test. Most activity can still be observed from 5 s to 8 s but apart from that, there is only little resemblance between the analysis results for the reference and the recorded signal.

Metrics

Based on the correlograms that are provided by the binaural hearing model both for the reference $\Psi_{\text{ref}}(k, m)$ and for the recorded signals $\Psi_{\text{rec}}(k, m)$, different metrics can be derived which should (at least partially) quantify the impact of spatial fidelity on the perceived audio quality. All the metrics in this contribution are calculated on the basis of the correlograms after weighting and averaging $\Psi(k, m)$ (c.f., the visualizations in Figures 1 to 3).

A degradation in signal quality when comparing reference and recorded signal does lead to a change in the correlogram. Thus, the basis for any metric is a difference $d(k, m)$ between the two correlograms.

$$d(k, m) = \Psi_{\text{rec}}(k, m) - \Psi_{\text{ref}}(k, m) \tag{1}$$

This definition would give positive values for components in the correlogram that are added by the reproduction system and negative values for missing components. For many application scenarios, it does not matter which type of change is present but any change from the reference is considered problematic. Hence, the absolute value of the difference is calculated as well.

$$d_{\text{abs}}(k, m) = \left| \Psi_{\text{rec}}(k, m) - \Psi_{\text{ref}}(k, m) \right| \tag{2}$$

For both calculations of the distance between the correlograms, different metrics can be derived from the statistics of the particular distance. Here, five different variants are considered (with Ndenoting the total number of

time-lateralization points in the correlogram). The same statistical quantities can be determined for the absolute value of the difference - they are omitted for brevity here.

- Mean value

$$\bar{d} = \frac{1}{N} \cdot \sum_{1}^{N_k} \sum_{1}^{N_m} d(k,m) \tag{3}$$

- Median value

$$d_{\text{median}} \tag{4}$$

- 5^{th}, 90^{th} and 95^{th} percentile

$$d_{\text{P05}}, d_{\text{P90}}, d_{\text{P95}} \tag{5}$$

These variants were chosen as both the mean as well as the median value are usually a good indicator for the overall difference between the correlograms while the different percentiles can serve as a characterization of the statistical spread of the difference.

Simulation Results

The ten different metrics were determined for the stimuli used in the listening test. The Pearson correlation coefficient between the individual metrics and the auditory results was calculated. The results are given in Table 1.

Metric	Correlation coefficient
$\bar{d}$	0.040
d_{median}	-0.081
d_{P05}	-0.321
d_{P90}	0.140
d_{P95}	0.192
$\bar{d}_{\text{abs}}$	0.160
$d_{\text{absmedian}}$	0.020
d_{absP05}	-0.153
d_{absP90}	0.173
d_{absP95}	0.212

Table 1: Correlations between the auditory results and the ten different metrics

It is obvious that none of the metrics alone is remarkably correlated with the results of the auditory test. Two scatter plots are given in Figures 4 and 5 to visualize the weak dependance between the tested metrics and the results of the listening test. In all scatter plots, the auditory results are given on the abscissa while the ordinate is used for the metric values or (later on) for the regression results.

The individual metric with the largest correlation coefficient is visualized in Figure 4: the 5^{th} percentile of the difference between the two correlograms. As could be expected from the numerical value (0.321), no usable connection between the auditory results and the metric can be observed. There is a slight trend towards higher auditory results for lower values of d_{P05} but nothing really meaningful can be deduced from this.

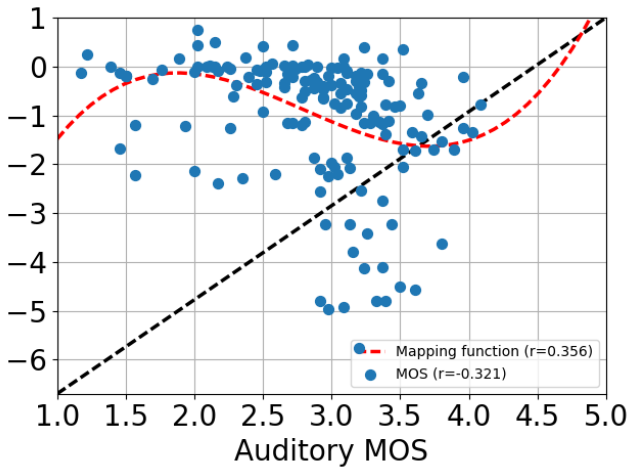

Figure 4: Scatter plot for d_{P05}

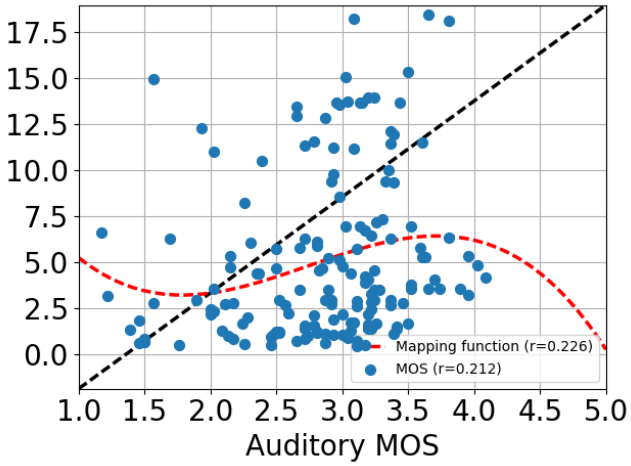

Figure 5: Scatter plot for d_{absP95}

The same interpretation also holds for the second-best metric in Figure 5: the 95^{th} percentile of the absolute value of the difference between the correlograms. Again, no clear connection between metric and auditory results but only a small trend can be seen: better auditory results for higher values of d_{absP95}.

Despite the poor performance of the individual metrics, a final judgment is not yet possible. Instrumental assessment is practically always based on the combination of different metrics by means of a trained regression. There is a wide range of possibilities for this regression: While straightforward multi-dimensional linear regressions are still used in certain applications, more complicated problems are nowadays usually approached by machine learning approaches, e.g., neural networks or decision tree learning. One type of decision tree learning, a Random Forest Regressor [9], is used here for testing if a combination of several metrics can lead to better performance. In a first experiment, all the available audio samples (with their respective auditory results) are used for training. Figure 6 shows the resulting scatter plot of instrumental assessment against auditory result.

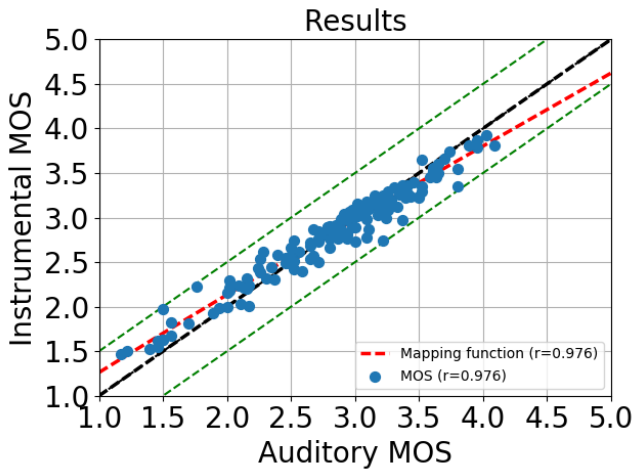

Figure 6: Scatter plot for training on all samples

The results are very good: The correlation is high and no major outliers occur. Given that the amount of available training data is fairly small, this is not really surprising – with enough input features, powerful machine learning algorithms are usually capable of memorizing the training data even if the features themselves have no or only little relation to the target variable. The regression here was even deliberately parametrized to trade a bit of performance in this unrealistic case for better generalization.

A fairer evaluation of the capabilities of the regression (and thereby the devised metrics) needs disjunct data sets for training and validation. Thus, 81 of the available samples were used for training in a second experiment. Only the results for the remaining 80 validation samples are depicted in Figure 7.

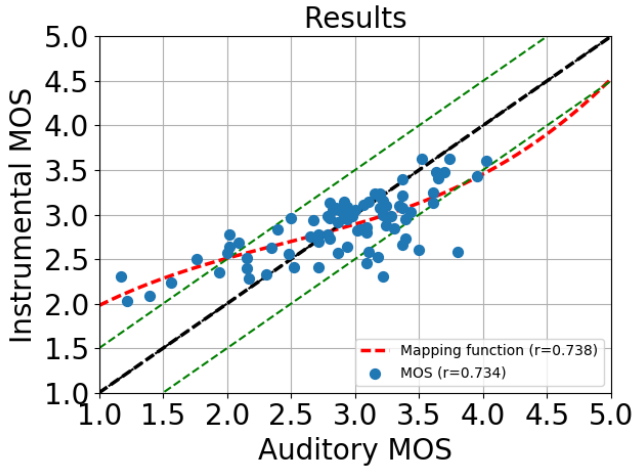

Figure 7: Scatter plot for training on half of the samples

It can be seen that the relation between the instrumental assessment and the auditory results is more pronounced than for the individual metrics (cf., Figures 4 and 5). However, the correlation is still only 0.734, there are some strong outliers and the instrumental assessment exhibits a clear tendency to assign the average value of all samples (2.86) to each individual sample.

Conclusions

An approach for instrumental assessment of audio quality by spatial analysis was evaluated. A binaural hearing model was used to calculate correlograms for the reference and for the recorded signal. Based on the difference between the two, ten different metrics were devised for quantifying the impact of spatial features on the perceived quality. The approach was evaluated using the audio signals and the auditory results of a recently conducted listening test. The individual metrics are only marginally correlated with the auditory results. Combining the metrics by a Random Forest Regressor improves the performance but it is still far from satisfactory.

It was shown by this evaluation that spatial properties alone are not sufficient for the quantification of the overall audio quality. Future work will focus on other signal properties to find more meaningful differences between the reference and the recorded signals.

References

[1] Magnus Schäfer, Mohammad Bahram, and Peter Vary. An extension of the PEAQ measure by a binaural hearing model. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, May 2013.

[2] Magnus Schäfer. *Multi Channel Audio Processing: Enhancement, Compression and Evaluation of Quality*. PhD thesis, RWTH Aachen, Aachen, 2014.

[3] ITU-R Recommendation BS.1387-1. *Method for objective measurements of perceived audio quality*, November 2001.

[4] Jan Reimes, André Fiebig, Thomas Deutsch, and Michael Oehler. Comparison of Auditory Testing Environments for Car Audio Systems. In *Fortschritte der Akustik - DAGA 2017*. DEGA e.V., Berlin, 2017.

[5] Werner Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80(6):1608–1622, 1986.

[6] Werner Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front. *The Journal of the Acoustical Society of America*, 80(6):1623–1630, 1986.

[7] Magnus Schäfer, Mohammad Bahram, and Peter Vary. Improved Binaural Model for Localization of Multiple Sources. In *10. ITG Symposium on Speech Communication*, Braunschweig, Germany, Sept 2012.

[8] Roy Patterson, Ian Nimmo-Smith, John Holdsworth, and Peter Rice. An Efficient Auditory Filterbank Based on the Gammatone Function. Technical report, IOC Speech Group on Auditory Modelling at RSRE, Dec 1987.

[9] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.