Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

Predicting perceived off-center sound degradation in surround loudspeaker setups for various multichannel microphone techniques

Nils Peters[1], Bruno L. Giordano[1], Sungyoung Kim[1], Jonas Braasch[2], and Stephen McAdams[1]

[1] *McGill University, Schulich School of Music and CIRMMT, 555 Sherbrooke Street West, Montréal, PQ, H3A 1E3, Canada*

[2] *Rensselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, NY, 12180-3590, US*

Correspondence should be addressed to Nils Peters (nils.peters@mcgill.ca)

ABSTRACT

Multiple listening tests were conducted to examine the influence of microphone techniques on the quality of sound reproduction. Generally, testing focuses on the central listening position (CLP), and neglects off-center listening positions. Exploratory tests focusing on the degradation in sound quality at off-center listening positions were presented at the 123rd AES Convention. Results showed that the recording technique does influence the degree of sound degradation at off-center positions. This paper focuses on the analysis of the binaural re-recording at the different listening positions in order to interpret the results of the previous listening tests. Multiple linear regression is used to create a predictive model which accounts for 85% of the variance in the behavioral data. The primary successful predictors were spectral and the secondary predictors were spatial in nature.

1. INTRODUCTION

Surround audio reproduction is known as a non-democratic reproduction technique because only the listener in the central listening position (CLP), also known as the "sweet spot", perceives the best audio quality, whereas off-center listening positions are generally considered as non-ideal. When a surround recording is presented to a larger audience this problem is critical, because most listeners will be located at off-center positions and exposed to a degraded sound image. So far, the influence of the listening position was primarily studied in terms of localiza-

tion errors, either based on binaural auditory models (e.g. [10], [19]) or through psychometric listening tests (e.g. [15], [1]). The perceptual evaluation of other quality attributes across different listening positions is of increasing interest to researchers ([17], [23]). Usually, sound quality of spatial audio reproduction systems is primary evaluated from the CLP, and a number of perceptive models have been created in order to predict the overall reproduction sound quality for different purposes. A review can be found in [2]. The studies [21] and [22] examined the perceptual attributes important to listeners of spatial audio reproduction. The "Basic Audio Quality", one of the defined global judgments for sound quality (BAQ), was evaluated through the controlled degradation of the spatial and timbral attributes of 5.1 surround audio material. A regression model was developed that shows the contribution of these attributes to the BAQ. It was concluded that timbre has a fairly strong weight on the BAQ (ca. 70%). The contribution of the two spatial attributes "frontal spatial fidelity" and "surround spatial fidelity" differed across listener groups, but can be considered as ca. 30% in total. In these experiments the subjects were placed at the CLP of an ITU 5.1 loudspeaker arrangement. One could hypothesize that listeners judge the degree of sound degradation at off-center listening positions with a somehow similar ratio of spatial and timbral fidelities.

To predict listener preference from objective physical measures is a relatively unexplored practice and one of the ongoing challenges in research on sound quality. Approaches differ in terms of several aspects:

1. The methodology for the collection of behavioral data.
2. The extracted physical measures.
3. The extraction process.
4. The applied statistical methods to create the predictive model.

In [11], the authors used a "Ridge Regression Model" to predict the MUSHRA-rated ([13]) preferences for "Frontal Spatial Quality" and "Surround Spatial Fidelity" for different spectrally and spatially degraded (bandwidth limitation and down-mixing) five-channel program items. The spectral features were measured from a mono mix-down, where all loudspeaker channels were summed together. For measuring spatial features binaural recordings were synthesized using convolution of the loudspeaker signals with KEMAR dummy-head HRTFs from a database for different head positions. From all 22 extracted features, measurements based on the Interaural cross-correlation coefficient (IACC, see section 3.2) as well as the spectral features "Centroid of the spectral coherence" and "Spectral Roll-Off" had the most impact on the rating. The predictive models showed a regression coefficients of $R = 0.91$ for the "Frontal Spatial Quality" and $R = 0.95$ for "Surround Spatial Fidelity" with the subjective ratings.

In a pairwise-comparison listening experiment [8], auditory attributes were retrieved from pop music and classical excerpts, which were presented through eight different multichannel reproduction formats (from mono to 3/2 surround). The ratings of seven of these auditory attributes were correlated with seven spatial and spectral measures. The success of the correlation varied between the musical genres in general, but the spatial measures based on "IACC" and "Lateral Fraction" accounted well for the variance in the spatial auditory attributes "Width" and "Spaciousness". In contrast, no significant correlation was found between any auditory attribute and the spectral measures "Spectral Centroid" and "Sharpness".

Kim et al. analyzed in [14] the overall preference choice between surround microphone techniques in several piano recordings. Eighteen features were extracted from a dummy-head re-recording captured at the CLP of a 5.0 surround loudspeaker set-up. A stepwise multiple regression model revealed that the measures "Ear Signal Incoherence" and "Side Bass Ratio" predicted the preference ratings of two separate groups of listeners reasonably well ($R^2 = 0.86$, $R^2 = 0.83$).

A more recent paper focused on the perceived sound quality of multichannel compression codecs [7]. The perceived mean opinion scale (MOS) of 11 multichannel audio excerpts processed by 11 different multi-channel audio compression codecs was predicted. For this purpose, a single layer feed forward neural network system was applied and trained with half of the 121 rated audio excerpts. Based on five spectral and two spatial measures, the model pre-

dicted the MOS of the rest of the audio excerpts. The correlation coefficient between measured and predicted MOS was 0.77.

2. METHODOLOGY

In a previous study [17], the authors addressed the question of whether different microphone techniques affect the size of the sweet spot in a 5.0 multichannel sound system. Two different 5.0 multichannel musical experts (see Table 1), each simultaneously recorded with three different microphone techniques (see Table 2) in the typical "F-B" fashion[1], were presented in two different large rooms through a 5.0 multichannel sound system.

EXC	Description
BACH	J.S. Bach: "Variation 13", Goldberg Variationen (BWV 988). Solo piano performance
MOZART	W.A. Mozart "Maurische Trauermusik", (KV 477) c-minor. Symphony performance

Table 1: Presented musical excerpts (EXC)

RT	Description
OMNI	in BACH: Polyhymnia Pentagon in MOZART: Decca Tree + Hamasaki-Square
OCT	Optimized Cardioid Triangle
AMBISONICS	Soundfield MKV + SP451 Processor

Table 2: Recording techniques (RT)

Binaural stimuli (BS) were recorded with a B&K Head-And-Torso-Simulator (HATS) at twelve different listening positions, including the CLP. In the following headphone-based listening experiment, 19 subjects compared these binaurally captured soundfields from different off-center positions with the soundfield captured at the CLP. In this pairwise comparison task, participants rated sound degradations moving a slider along a continuous scale from 0 (total degradation) to 100 (no degradation). "Sound Degradation Maps" were given by the average ratings (see [17] for details). Interestingly, the perceived sound degradation differed not only across the three recording techniques, but also significantly between the musical excerpts for one of the tested rooms (see Figure 5).

Here we extend the analyses of behavioral data reported in the first experiment of [17], exploring the relationship between acoustical features of the stimuli and perceived sound degradation.

Figure 1 shows the relation of analyzable audio material to the gathered behavioral data. Since subjects were exposed to the binaural stimuli (BS) it seems important to search for correlations in this direct relation. To understand the effect the RT has on the sound degradation, of particular interest is the "Indirect Relation" between physical measures from the original 5.0 multichannel surround recordings in combination with extracted features from Binaural Room Impulse Responses (BRIR) from each of the five loudspeakers for the different listening positions. If, according to signal theory, a binaural stimulus is a result of the original 5.0 channel surround recording being convolved with the BRIR of each loudspeaker, the physical features measured in the BS must consequently rely on measures from the BRIR

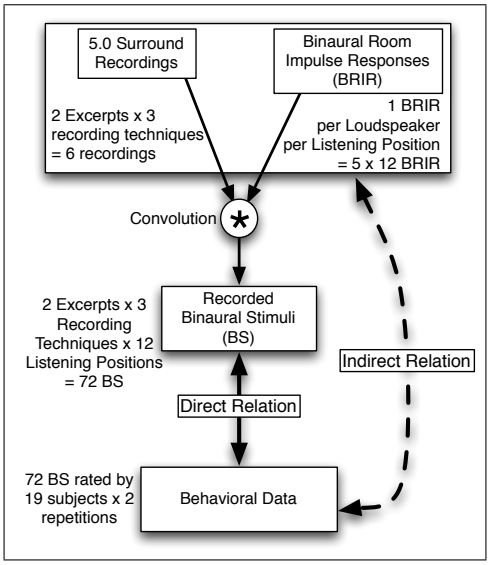


Fig. 1: Relations between audio material and behavioral data.

[1] F-B: To recreate the impression a listener has in a concert, the three front channels are used to recreate the instrument sounds coming from the stage, whereas the rear channels contain mainly ambient sounds and room response.

and the original 5.0 multichannel surround recordings. Since it is not clear how this "Indirect Relation" can be modeled perceptually, here we focus on the relation between the BS and behavioral data. The understanding of this relationship will also provide knowledge that fosters the understanding of the more complex dependencies.

3. FEATURE EXTRACTION

The features which were extracted from the BS were partly chosen based on the subjects' post experimental responses regarding their rating strategies.
In order to account for the time-varying properties of the acoustical measures, BS were analyzed across their duration (7 s ca.) using a sliding window of 50 ms and a hop-size of 25 ms. The final descriptors were extracted from the different time-varying acoustical features using the following statistical functions:

Mean value of the time series `x`: (`xMean`).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$$

Maximal value of the time series `x`: (`xMax`).

$$x_{Max} = \max(x)$$

Ratio between the maximal value and the mean the time series `x`: (`xRatio`).

$$x_R = \frac{x_{Max}}{\bar{x}}$$

Unbiased standard deviation of elements in the time series `x`: (`xStd`).

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

3.1. Spectral Features

The perception of spectral cues is considered to be primarily a monaural process [5]. Nevertheless, as shown in [6], perceived timbral modifications through the presence of strong room reflections can be suppressed by binaural mechanisms. The following operators symbolize the different functions used to measure both binaural and monaural spectral properties:

`L,R`: Left and right ear, separately.
`Diff`: Difference of the spectral measures between left and right ear.
`Avg`: Average of the left and right ear (one channel)

3.1.1. Spectral Centroid

The Spectral Centroid C_f, or spectral center of gravity, is closely related to perceived brightness. It is computed from the magnitude $M_j[n]$ of the FFT spectrum for each frame f (see Equation 1), where n = frequency bin number and N = Nyquist frequency:

$$C_f = \frac{\sum_{n=1}^{N/2} M_f[n]n}{\sum_{n=1}^{N/2} M_f[n]} \quad (1)$$

3.1.2. Sharpness

Sharpness was computed from the loudness specific to each of the critical bands, as described by Zwicker and Fastl in ([25]):

$$S = 0.11\frac{\int_0^{24Bark} N^{'}(z)g(z)zdz}{\int_0^{24Bark} N^{'}(z)dz}acum \quad (2)$$

with $N^{'}(z)$ = the specific loudness and $g(z)$ = a weighting factor depending on the critical-band rate.

3.1.3. Spectral Roll-Off

The Spectral Roll-Off R_f represents the frequency below which 95% of the frame's signal energy exists. It is correlated to the harmonic cutting frequency.

$$\sum_{n=1}^{R_f} M_f[n] = 0.95\sum_{n=1}^{N/2} M_f[n] \quad (3)$$

3.1.4. Spectral Flux

The Spectral Flux SF measures how quickly the spectrum changes. It calculates the difference between two consecutive frames M_f and $M_{[f-1]}$ for each frequency bin n.

$$SF = \sum_{n=1}^{N}(M_f[n] - M_{[f-1]}[n])^2 \quad (4)$$

3.1.5. Energy Features

Energetic measures were computed for the entire spectrum, as well as for the different octave bands with center frequency from 64 to 8000 Hz. Instead of using a finer frequency resolution (e.g. 3rd octave bands or less), these relatively broad spectral bands were chosen in this exploratory state.

3.2. Interaural Features

Since the stimuli were binaurally recorded several features related to the interaural properties were extracted. The Interaural cross-correlation coefficient (IACC) is widely used in concert hall acoustics where it is known to be a predictor of apparent source width (ASW), spaciousness and listening envelopment (LEV) (see e.g. [16]). The classic IACC is calculated from a binaural room impulse response (BRIR) and is the absolute maximum of Φ in equation 5, calculated within a range of $\tau = \pm 1$ ms. The operators B_L and B_R represent the left and the right ear signal.

$$\Phi_{lr}(\tau) = \frac{\int_{T_0}^{T} B_L(t) \cdot B_R(t+\tau)dt}{\sqrt{\int_{T_0}^{T} B_L^2(t)dt \cdot \int_{T_0}^{T} B_R^2(t)dt}} \qquad (5)$$

In order to express the effect of early and late reflections on the perceived ASW and LEV, several IACC sub-measures were developed which use different upper and lower temporal integration boundaries T. It is an ongoing discussion whether room acoustic measures, such as the IACC, are relevant for evaluating audio reproduction, especially because many of the concert hall measures are based on the analysis of (binaural) room impulse responses (see e.g. [24]). Differently from the case of impulse responses, with the analysis of musical signals it is not possible to apply the temporal integration boundaries T_0 and T. Hence, for this study the IACC-based features are calculated from the binaural signals within the sliding window duration of 50 ms and a hop-size of 25 ms. To represent the variations of the spatial features in time, the `Mean`, `Max`, `Ratio` and `Std` values are calculated from all retrieved IACC maxima.

3.2.1. Octave band IACC

Besides the broadband IACC, which gives the IACC over the entire spectrum, an IACC was also extracted in eight separate octave bands with center frequencies between 64 Hz and 8000 Hz.

3.2.2. Modified IACC

In [8], a modified version of the broadband IACC was proposed. In order to simulate the extraction of the envelope carried out within the auditory system, the binaural signal is treated with a half-wave rectification followed by a 1 kHz low-pass filtering before the classical IACC calculation is applied. As reported in [16], this modified IACC represents the ASW for narrow-band sounds better than the conventional IACC.

3.2.3. Binaural Quality Index

The Binaural Quality Index (BQI) is typically used as a measure in concert hall acoustics. According to [4], the BQI *"is one of the most effective indicators of the acoustical quality of concert halls"*. It is defined as the average IACC value for early reflections up to 80 ms in the 500 Hz, 1000 Hz and 2000 Hz octave bands. According to [4], for good concert halls the BQI was found to be above 0.5.

$$BQI = 1 - IACC_{[500,1000,2000]} \qquad (6)$$

3.3. Feature Rescaling

Since each off-center recording was judged against the reference recording taken at the CLP in the listening experiment, the extracted features had to be rescaled in order to indicate this relation. Two rescaling methods were used to express the degradation of the acoustical features in relation to the reference.

As shown in equation 7 the difference between the acoustical feature X_i extracted from the BS at listening position i and X_{Ref} taken at the reference (CLP) is multiplied by X_{Ref}. This rescaling method was proposed in [7] to express the distortion of interaural measures for the purpose of comparing audio compression algorithms.

The second rescaling method (equation 8) calculates the absolute value of the previously described term.

$$X_i^{r_1} = X_{Ref} \cdot (X_{Ref} - X_i) \qquad (7)$$

$$X_i^{r_2} = |X_i^{r_1}| \qquad (8)$$

The rescaling methods differ in the fact that equation 8 calculates the strength of the difference, whereby equation 7 also takes the direction of the difference into account. Both rescaling methods are necessary because we do not know *a priori* whether

the strength of the difference matters or whether the direction of the difference is important in the subject's rating strategy.

4. RESULTS

4.1. Stepwise Linear Regression

A stepwise linear regression was performed taking the ratings of 72 BS into account (2 musical excerpts (EXC) x 3 recording techniques (RT) x 12 listening positions (POS)). The anchor stimuli providing the low-quality reference within the experiment, were not considered in this analysis, since the artificial manipulation used to generate is not normally encountered in real listening contexts.

A linear regression model with four predictors was created which accounted for 84% of the variance in the behavioral data ($R^2 = 0.84$, $R^2_{adj.} = 0.83$, $p < 0.001$). By neglecting four outliers which are outside of the model's 95% confidence interval, the goodness of fit increased to $R^2 = 0.90$, $R^2_{adj.} = 0.89$. The four successful predictors and the phases of the stepwise selection process are displayed in Table 3. The feature `CenStdAvg`[2] was suggested by the stepwise procedure as a fifth predictor. It was decided to keep the model with four predictors because the fifth predictor would improve the fitness only marginally at the cost of increasing the model's complexity. As shown in Table 4 the correlation between these four predictors was found to be consistently low. Figure 2 shows the created model. An analysis of the model's residuals confirmed that the behavioral data of both musical excerpts are similarly well predicted. A small difference across the three RTs can be observed: Behavioral data related to Ambisonics are better predicted than the data of the two other RTs.

Draft No.	Accum. $R^2_{adj.}$	Predictor	Partial R	Standard. coeff. B
1.	.67	`RMS8000Avg`	.83	.65
2.	.74	`RMS250Avg`	.59	.31
3.	.80	`|RMS8000Diff|`	-.49	-.23
4.	.84	`IACC1000Std`	-.44	-.20

Table 3: Stepwise linear regression model.
RMS8000Avg: the energy in the 8000 Hz octave-band for the averaged ear signals; RMS250Avg: the energy in the 250 Hz octave-band for the averaged ear signals; |RMS8000Diff|: the absolute difference of the energy in the 8000 Hz octave-band between the ears; IACC1000Std: the standard deviation of the IACC-value in the 1000 Hz octave-band.

[2]CenStdAvg: the standard deviation of the spectral centroid of the averaged ear signals.

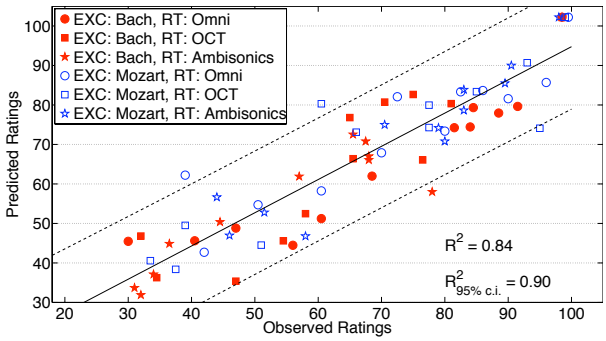


Fig. 2: Results of multiple regression analysis.
Filled red markers symbolize the BACH excerpt, unfilled blue markers symbolize the MOZART excerpt. Dashed lines show the 95% confidence interval about the regression line.

	`RMS250 Avg`	`|RMS8000 Diff|`	`IACC1000 Std`
`RMS8000Avg`	.26	-.23	-.18
`RMS250Avg`	1.	.06	-.08
`|RMS8000Diff|`		1.	.10

Table 4: Pearson Correlation between predictors.

4.2. Cluster Analysis

Despite the good prediction of the experimental data, the stepwise selection procedure tends to discard from the final regression model acoustical variables that are otherwise strongly correlated with the behavioral data. We adopted a data reduction method guided by a cluster analysis to overcome this limitation [12].

The initial stage of this procedure requires computing a measure of the distance between acoustical variables, given by the absolute value of their Spearman rank correlation (the rank correlation is independent of the particular nonlinear monotone relations between variables). The distances are then analyzed with a clustering method (agglomerative hierarchical cluster analysis, average linkage). Finally, each of the clusters of strongly correlated acoustical descriptors is merged independently into a single variable by means of Principal Component Analy-

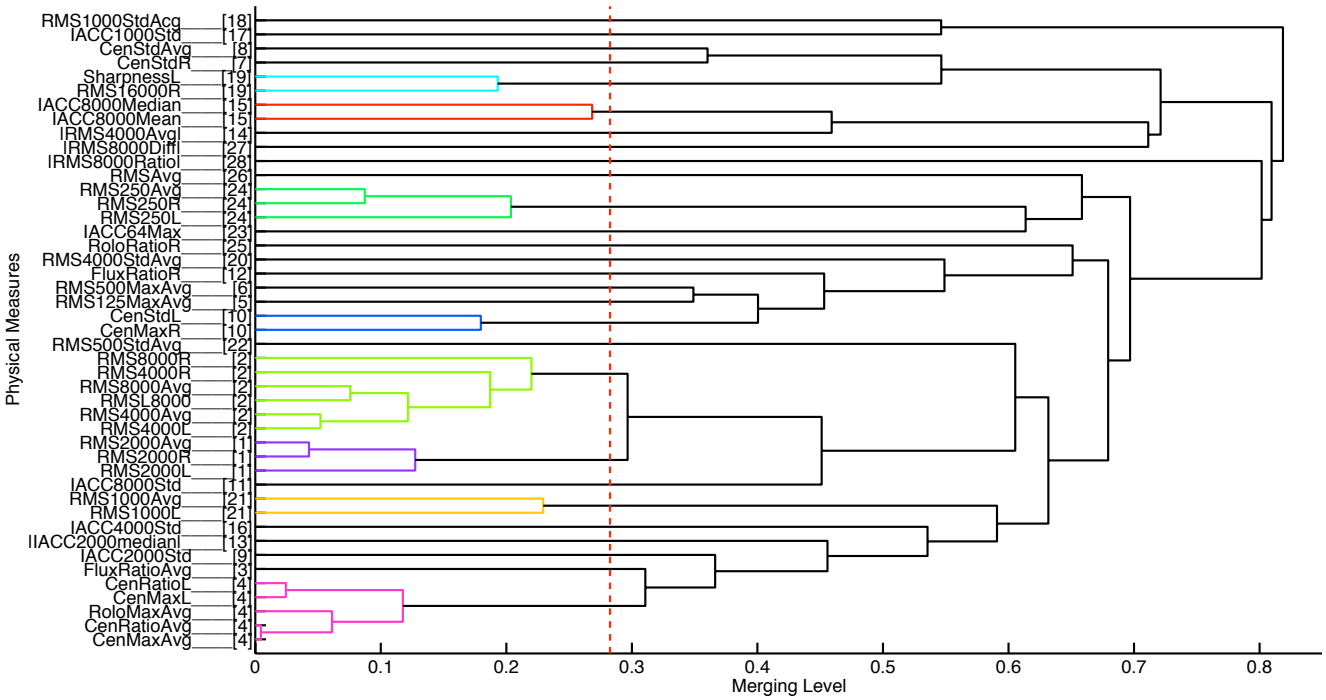


Fig. 3: Hierarchical cluster analysis of physical measures. The vertical red dashed line shows the level until the reduction process was performed to obtain the model shown Table 5. Cluster No. in hard brackets.

sis (PCA). The first Principal Component (PC) for each of the clusters is retained as the final reduced variable. Starting from the condition where each acoustical descriptor is in an isolated cluster, the number of clusters can be progressively decreased, thus yielding an increasingly lower number of reduced variables. As the number of clusters is decreased, one observes a decrease in the correlation between reduced acoustical descriptors, and a decrease in the extent to which the original acoustical descriptors are well accounted for through the remained clusters: the final number of clusters and of acoustical descriptors is then selected trading off these two factors.

Figure 3 shows the cluster analysis that guided the data reduction process. The data-reduction process was carried considering only those 45 acoustical features moderately-to-strongly correlated with the behavioral data ($|R_s| > 0.40$). The final number of clusters was 28 (see vertical red dashed line in Figure 3). Due to this process, the maximum absolute correlation across the cluster items decreased from $R = 0.99$ to $R = 0.78$ whereas the clusters accounted well for the reduced acoustical variables (minimum absolute correlation between acoustical variables and respective clusters = 0.81).

In the following, a stepwise linear regression was carried out with the PCs of descriptors as predictors. The final model is shown in Figure 4. The final model included five PCs, and accounted for 85% of the variance in the behavioral data ($R^2 = 0.85$,

Draft No.	Accum. $R^2_{adj.}$	Predictor	Partial R	Standard. coeff. B
1.	.68	Cluster No.2	-.81	-.61
2.	.73	Cluster No.8	-.36	-.17
3.	.77	Cluster No.17	-.39	-.17
4.	.80	Cluster No.24	-.48	-.23
5.	.84	Cluster No.27	-.47	-.22

Table 5: Stepwise linear regression model of the clustered measures. Cluster No. refers to the cluster index in Figure 3.

$R^2_{adj.} = 0.84$, $p < 0.001$). The included clusters are shown in Table 5 and their Spearman correlation coefficients in Table 6. The physical measures that were part of the model created in section 4.1 are also included in the selected clusters for creating this model. An interesting contribution to the model shows the cluster No.2. Beside the physical measure `RMS8000Avg`, this cluster contains energy features of the octave bands of 4000 Hz, and 8000 Hz. Furthermore, the cluster No.24 was created from the measures `RMS250Avg`[3], `RMS250L` and `RMS250R`. It seems reasonable that these measures are strongly correlated since in this frequency range no head shadowing effect is present. The other successful clusters are similar to the previous predictive model, but were selected in a different order. For example, the cluster No.8 which contains `CenStdAvg`[4] was chosen as the second entry. In the model from section 4.1, `CenStdAvg` was proposed as the fifth predictor without having a strong effect on the model. According to the model goodness of fit, it can be concluded that the model from section 4.1 and the cluster-based model have practically the same performance. Yet, the fact that the cluster-based model respects the degree of correlation across the physical measures makes this model more meaningful. Apart from this, it can also be observed what physical measures are similar to each other even if they do not appear in the final predictive model: For example, the cluster No. 4 contain measures related to the Spectral Centroid and the Spectral Roll-off.

	Cluster No. 8	Cluster No. 17	Cluster No. 24	Cluster No. 27
Cluster No.2	.39	.18	.32	.22
Cluster No.8	1.	.18	.31	.18
Cluster No.17		1.	.12	.10
Cluster No.24			1.	-.06

Table 6: Pearson Correlation between clusters.

4.3. Predicting the RT with the least perceived sound degradation

As shown in Figure 5, differences were found in the perceived sound degradation across the different recording techniques (RT) per position (POS). To explore whether the extracted physical measures can predict the RT with the least perceived sound degradation per position, a Spearman Rank Correlation between the three RTs and all physical measures was performed. Table 7 shows the physical measures which lead to $|R_s|$ correlation values higher than 0.6 separately for each musical excerpt. The calculated correlation values do not correspond sufficiently across the EXCs. Especially surprising is the correlation of `RMSMaxL`[5] which gives a strong positive value for EXC Bach, but a strong negative value for EXC Mozart. Therefore it must be concluded that either the extracted physical measures do not completely cover all relevant aspects to

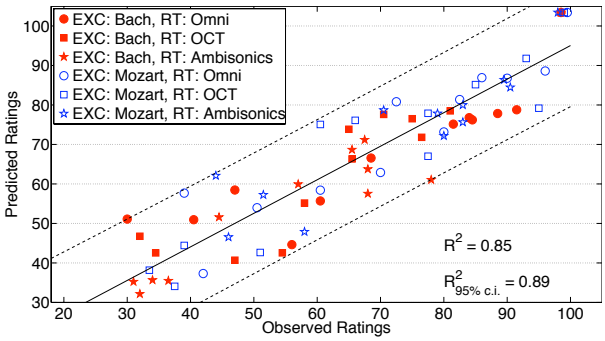


Fig. 4: Results of multiple regression analysis of the clustered features.
Filled red markers symbolize the BACH excerpt, unfilled blue markers symbolize the MOZART excerpt. Dashed lines shows the 95% confidence interval about the regression line.

Physical Measure	EXC Bach	EXC Mozart
`IACC250Max`	**.63**	.36
`|IACC500Mean|`	-.13	**-.63**
`|IACC4000Mean|`	**-.64**	-.18
`IACC8000Mean`	**-.64**	-.32
`RMSMaxL`	**.63**	-.55
`FluxMeanL`	**.64**	-.36
`FluxMaxAvg`	-.36	**-.68**
`FluxRatioAvg`	-.50	**-.63**
`CenStdAvg`	-.41	**-.68**

Table 7: Spearman Rank Correlation between RT and physical measures.

[3]RMS250Avg: The energy in the 250 Hz octave band of the averaged ear signals.

[4]CenStdAvg: the standard deviation of the spectral centroid of the averaged ear signals.

[5]RMSMaxL: the maximum dB value at the left ear signal.

describe the perceived differences, and/or that the prediction of the RT with the least perceived sound degradation is a more complex (multidimensional) affair. However, the only measures that create relatively similar strong correlations for both EXCs are `FluxRatioAvg`[6] and `CenStdAvg`[7]: The RT with the lowest measured values in `CenStdAvg` or as well in `FluxRatioAvg` was rated as the RT with the least perceived sound degradation on 7 of 11 off-center positions. This is a probability of ca. 64%, which is higher than chance (33.3%).

5. INTERPRETATION

The cluster-based regression model uses primary spectral measures to predict the behavioral data of 19 subjects. The most successful predictor is the spectral energy in the mid and high frequencies (Cluster No. 2). The lower the energy in these frequency bands (related to the CLP, see section 3.3), the stronger the perceived sound degradation. The reason for the loss in mid and high frequencies might be due to the natural loudspeaker directivity in this frequency range. Since all loudspeakers are directed to the CLP, all spectral information is present, but the loss of mid and high frequencies becomes stronger the more the listener is located off-center. Mid frequencies might reach the listener through early reflections, but the higher the frequency, the more energy is absorbed by the air and by room materials. The predictor `|RMS8000Diff|`[8] can be interpreted in a similar manner. For most off-center listening positions, one ear is closer to the CLP than the other ear. This energy difference, heightened through the head shadowing effect, seems to matter in the 8000 Hz octave band.

The energy of the 250 Hz octave band, represented through the predictors `RMS250Avg`, `RMS250L` and `RMS250R`, also appears crucial for the perceived sound degradation. Further testing is needed to interpret this finding. Time-of-flight differences between the loudspeaker signals at off-center listening positions might cause perceptible comb filter effects in this frequency range, while additional interference between direct and reflected sounds might exacerbate this.

The only successful interaural feature is `IACC1000Std`[9]. Since `IACC1000Std` is expressed through a negative standardized coefficient B (see Table 3), the perceived sound quality increases if `IACC1000Std` decreases. In other words, subjects perceived a fluctuation of the 1000Hz-IACC measures as a degrading aspect. It is known that frequencies up to ca. 2000 Hz contribute to the spatial impression, especially components around 600 Hz [20]. A fluctuation in the 1000Hz-IACC measures could therefore lead to a distortion in the perception of the spatial impression.

6. DISCUSSION

A regression model was created which predicts the ratings of perceived sound degradation reasonably well. As reviewed in the Introduction, the practice of correlating behavioral data with physical measures extracted from binaural re-recordings has been reported in previous sound quality studies. In these studies the behavioral data are usually gathered by exposing the subject to soundfields created by the loudspeaker, whereby the physical measures are obtained from binaural re-recordings. In other words, the predictive models are based on subjective and objective data retrieved from different sources. The difference in this study is that the binaural re-recordings are used as stimuli in the listening experiment as well as to obtain the physical measures.

This raises the issue concerning which approach might be more reliable. A listening experiment must be designed to allow real-time, double-blind, comparative and repeatable evaluations. Using *in situ* listening test methods would make it very difficult and almost impossible to follow these design criteria. Therefore the described method was used. The absence of head movements in fixed binaural recordings causes localization errors mainly in the median plane and in the region of the cone-of-confusion. A binaural room scanning system (BRS), which allows head movements through head tracking in the binaural reproduction system, reduces localization errors and increases out-of-head localization. However, the advantages of binaural head-tracking displays appear to diminish when room reflections are included in the

[6] FluxRatioAvg: the ratio between the maximum value and the mean value of the spectral flux for the averaged ear signals.

[7] CenStdAvg: the standard deviation of the spectral centroid of the averaged ear signals.

[8] |RMS8000Diff|: the absolute energy difference between the ears in the 8000 Hz octave band.

[9] IACC1000Std: the standard deviation of the IACC in the 1000 Hz octave band.

capturing process [3]. Since the static reproduction process was the same for all stimuli in the listening experiment, it can be assumed that the effect generates a constant bias for al the stimuli. Moreover, by using such a BRS system, the challenge of extracting physical measures from a binaural time-variant signal arises. Pfanzagl-Cardone and Höldrich presented recently a study [18] in which different surround microphone arrays were judged by: a) exposing the subjects to the loudspeaker-generated soundfield; and b) exposing the subject to a binaurally re-recorded version of the loudspeaker's soundfield presented over headphones. They found that the results of these two presentation methods differ less than expected and that *"the transformation process [through the binaural re-recording] may have 'amplified' the perceived differences between the surround-techniques under test. (Similar to a 'grayscaled' picture being converted into a 'black and white' image.)"*.

The second point of discussion refers to the contribution of the extracted features to the predictive models. The interaural measures did not show a strong correlation with the behavioral data. Comparable to the observations the authors of [9] made by measuring classical IACC values in concert halls, a large fluctuation of the adapted IACC measures on small spatial intervals are measured, even if no perceptual changes are perceived. Therefore, the adapted IACC measures fail to differentiate between listening positions.

A possible reason why timbral aspects matter the most might be attributed to the fact that the subjects of the listening experiment were sound recording students, educated through technical ear training methods. Since technical ear training methods usually focus on the sensitization to spectral cues, this suggests that training may bias the listener to give less attention to spatial aspects. However, distorted timbral perception may also result from non-ideal localization processes at off-center positions.

7. CONCLUSION & FUTURE WORK

The perceived sound degradation of two musical excerpts, each recorded using three different recording techniques at twelve different listening positions, was reasonably well predicted by a multiple regression model. The primary successful predictors are spectral and the secondary ones are based on spatial aspects. This order agrees somehow with the ratio of spatial and spectral aspects which has been found to describe the Basic Audio Quality (BAQ) in [21], [22].

An adaptation of the Binaural Quality Index, a successful measure in concert hall acoustics, but also other interaural measures, did not show strong relationships with the behavioral data and might not be applicable in evaluating off-center sound degradation in surround loudspeaker setups.

To generalize the findings, a cross-validation of this model needs to be performed, one that might also includes more psycho-acoustically motivated features, such as "Roughness", "Impulsiveness" or "Loudness". Furthermore, a separation of the frequency bands through auditory filters (e.g. Gamma-Tone Filter) might improve the model.

8. ACKNOWLEDGMENT

This work has been funded by the Canadian Natural Sciences and Engineering Research Council (NSERC). Thanks to Georgios Marentakis for assistance in this project.

9. REFERENCES

[1] E. Bates, G. Kearney, F. Boland, and D. Furlong. Localization accuracy of advanced spatialization techniques in small concert halls. In *153rd meeting of the Acoustical Society Of America*, 2007.

[2] S. Bech and N. Zacharov. *Perceptual Audio Evaluation—Theory, Method and Application.* John Wiley & Sons, Ltd, 2006.

[3] D. Begault, E. Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904 – 916, 2001.

[4] L. L. Beranek. *Concert Halls and Opera Houses: Music, Acoustics, and Architecture.* Springer-Verlag New York Inc., 2003.

[5] F. A. Bilsen. Pitch of noise signals: Evidence for a "central spectrum". *J. Acoust. Soc. Am.*, 61:150–161, 1977.

[6] M. Brüggen. Coloration and binaural decoloration in natural environments. *Acta Acoustica ACUSTICA*, 87:400–406, 2001.

[7] I. Y. Choi, S. B. Chon, B. G. Shinn-Cunningham, and K.-M. Sung. Prediction of perceived quality

in multi-channel audio compression coding systems. In *30th International Conference: Intelligent Audio Environments*, Saariselkä, Finland, 2007.

[8] S. Choisel and F. Wickelmaier. Relating auditory attributes of multichannel sound to preference and to physical parameters. In *120th AES Convention, Preprint 6684*, May 2006.

[9] D. de Vries, E. Hulsebos, and J. Baan. Spatial fluctuations in measures for spaciousness. *J. Acoust. Soc. Am.*, 110(2):947–954, 2001.

[10] M. Dewhirst, S. K. Zieliński, P. Jackson, and F. Rumsey. Objective assessement of spatial localisation attributes of surroung-sound reproduction systems. In *118th AES Convention, Preprint 6441*, Barcelona, Spain, May 2005.

[11] S. George, S. Zieliński, and F. Rumsey. Feature extraction for the prediction of multichannel spatial audio fidelity. *Audio, Speech, and Language Processing, IEEE Transactions on [see also Speech and Audio Processing, IEEE Transactions on]*, 14(6):1994–2005, Nov. 2006.

[12] B. L. Giordano, S. McAdams, and D. Rocchesso. Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Submitted*, 2008.

[13] International Telecommunication Union. *ITU-R BS.1116-1, Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems.* International Telecommunication Union, Geneva, Switzerland, 1997.

[14] S. Kim, W. L. Martens, A. Marui, and K. Walker. Predicting listener preferences for surround microphone technique through binaural signal analysis of loudspeaker-reproduced piano performances. In *121th AES Convention, Preprint 6919*, San Francisco, US, October 2006.

[15] G. Marentakis, N. Peters, and S. McAdams. Auditory resolution in virtual environments: Effects of spatialization algorithm, off-center listener positioning and speaker configuration (A). *J. Acoust. Soc. Am.*, 123(5):3798, 2008.

[16] R. Mason, T. Brookes, and F. Rumsey. Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli. *J. Acoust. Soc. Am.*, 117(3):1337–1350, March 2005.

[17] N. Peters, J. Braasch, and S. McAdams. Evaluating off-center sound degradation in surround loudspeaker setups for various multichannel microphone techniques. In *123th AES Convention, Preprint 7197*, New York, US, October 2007.

[18] E. Pfanzagl-Cardone and R. Höldrich. Frequency dependent signal correlation in surround- and stereo-microphone systems and the blumlein-pfanzagl-triple (BPT). In *124th AES Convention, Preprint 7476*, Amsterdam, The Netherlands, 2008.

[19] V. Pulkki. Microphone techniques and directional quality of sound reproduction. In *112th AES Convention, Preprint 5500*, Munich, Germany, 10–13 May 2002.

[20] J. Raatgever. *On the binaural processing of stimuli with different interaural phase relations.* PhD thesis, Delft University of Technology, 1980.

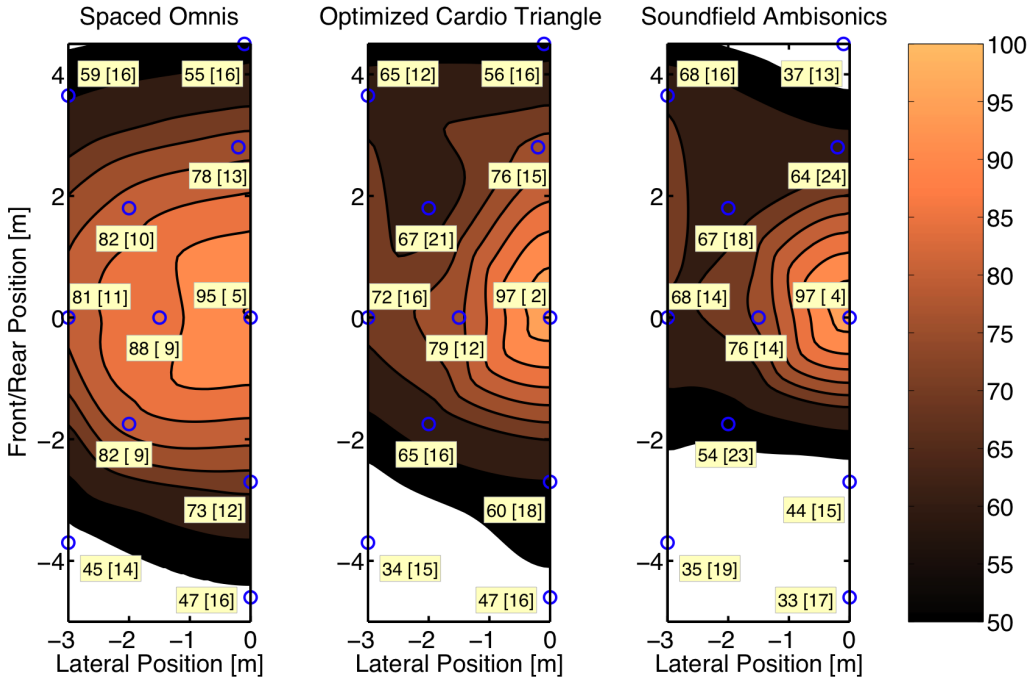
[21] F. Rumsey, S. Zieliński, and R. Kassier. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Audio Eng. Soc.*, 118(2):968–976, August 2005.

[22] F. Rumsey, S. Zieliński, R. Kassier, and S. Bech. Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences. *J. Acoust. Soc. Am.*, 117(6):3832–3840, June 2005.

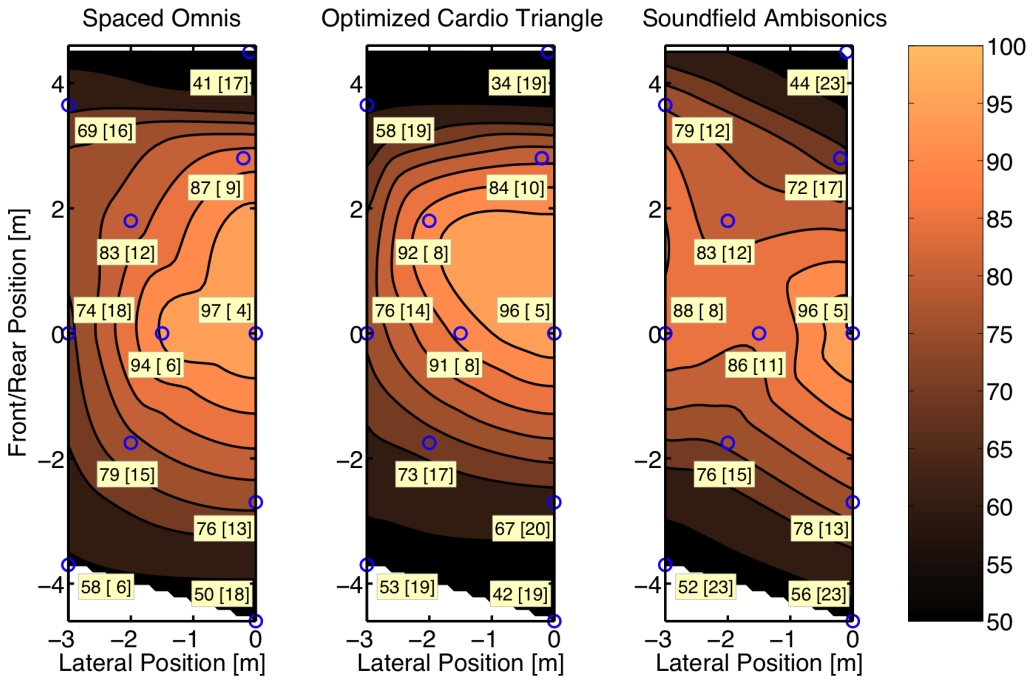
[23] F. J. Rumsey, S. Zieliński, P. J. Jackson, M. Dewhirst, R. Conetta, S. Bech, and D. Meares. Measuring perceived spatial quality changes in surround sound reproduction (A). *J. Acoust. Soc. Am.*, 123(5):2980, 2008.

[24] G. A. Soulodre. Can reproduced sound be evaluated using measures designed for concert halls? In *Workshop for Spatial Audio Sensory Evaluation Techniques*, www.surrey.ac.uk/soundrec/ias, Guildford, UK, April 2006.

[25] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models.* Springer, 3rd edition, 2007.



(a) EXC BACH



(b) EXC MOZART

Fig. 5: Mean ratings [standard deviation] for both musical excerpts. The tested listening positions are marked with blue circles. Contour-plots were created with cubic interpolation. Fig. reproduced from [17].