

# MULTICHANNEL LEVEL ALIGNMENT, PART IV: THE CORRELATION BETWEEN PHYSICAL MEASURES AND SUBJECTIVE LEVEL CALIBRATION

NICK ZACHAROV<sup>1</sup> AND SØREN BECH<sup>2</sup>

<sup>1</sup>Nokia Research Center, Speech and Audio Systems Laboratory, Finland.

[Nick.Zacharov@Nokia.com](mailto:Nick.Zacharov@Nokia.com)

<sup>2</sup>Bang & Olufsen A/S, Struer, Denmark

[sbe@bang-olufsen.dk](mailto:sbe@bang-olufsen.dk)

In an effort to establish a better understanding of multichannel level alignment, a correlation and regression analysis has been performed of objective metrics and subjective data collected and subsequently reported in parts I-III & V. This study considers subjective data acquired with 9 test signals with different set-ups considering loudspeaker placement, directivity, reproduction bandwidth and absolute reproduction level. Objective metrics considered include linear and weighted SPL and several loudness metrics. A constant specific loudness signal with a high pass filter at 500 Hz is found to provide the best prediction of subjective level alignment under a wide range of practical usage situations with a wide range of metrics.

## INTRODUCTION

This is the forth paper in a series considering level calibration of 5 channel sound reproduction systems. Numerous subjective experiments have been developed and performed as reported in parts I-III & V [1, 2, 3, 4]. Additionally, an overview of the multichannel level alignment issue can be found in [5]. The aim of the subjective tests performed to date has been to collect a database of subjective level alignment information under a wide range of practical domestic usage situation of 5 channel sound systems. The previous papers have attempted to assess which factors have perceptual relevance and which can be ignored. In practice a large number of factors can affect the level calibration of the multichannel sound system, as illustrated in figure 1, which show the ideal symmetric setup, asymmetric room acoustics and lastly the often encountered situation consisting of misplaced loudspeakers of different acoustics characteristics. The subjective experiments performed consider such setups within a practical range of limits under controlled conditions. The factors tested are presented in table 1.

The work of Aarts [6, 7, 8] and Bech [9] has suggested that the calibration signal plays an important role in the alignment of reproduction systems. Bech also suggested that the bandwidth of the signal might also influence the subjective level alignment in a critical fashion. Motivated by these studies, it became of interest to study a range of test signals (see table 2) in a level alignment task.

This paper examines the correlation between the subjective level calibration data collected in the previous experiments and objective measurements made in each experimental condition. The hypothesis under considering is whether a test signal/metric combination can be found that will provide

a perceptually valid level alignment compensating for differences associated with different source locations, distances, directivities, sensitivities and asymmetries associated with the room acoustics.

## 1. SUBJECTIVE EXPERIMENTS

In an effort to further study the subjective level alignment a set of experiments were designed to study various aspects of multichannel level alignment in isolation. The aspects considered include

- the source placement around the listener,
- the source distance,
- the setup symmetry,
- the source directivity,
- the source bandwidth,
- the absolute reproduction level,
- the listening room.

In an effort to better understand the influence of the test signal, a number of test signals were generated. A summary of the test signals can be found in table 2 and a full description and discussion of the signals can be found in [1].

### 1.1. General experimental design

In this section a brief summary of the experimental design and procedure is provided. The interested reader is referred to parts I-III & V for further details [1, 2, 3, 4]. The series of experiments has been performed in three different listening rooms at each of the test sites. Both the Nokia Research Center (NRC) and British Broadcasting Corporation (BBC) listening rooms are ITU-R BS.1116-1 [10] conformant and are described in [2] and [11] respectively. The Bang & Olufsen (B&O) listening room complies with the IEC 60268-13 standard [12]. The characteristics of the BBC listening room are illustrated in figure 4 and the reverberation characteristics are shown in figure 5.

Six listeners from each site were employed for each experiment. Members of the NRC and B&O permanent listening panels were employed, who are continuously and extensively trained in subjective listening and have normal hearing (i.e. less than 15 dB deviation from normal). All listeners at all sites have extensive experience in expert listening tests performed in accordance to ITU-R BS.1116-1 [10].

Each test system was calibrated by using the reference, centre channel. This channel's gain was adjusted to have an equal loudness for all signals in accordance to the Zwicker diffuse field method, as specified within ISO 532 [13]. A level of 20 Sones was found (see [2]) both to provide a comfortable listening condition and allow for sufficient headroom ( $> 6$  dB) in the reproduction system to avoid clipping. For experiment 4 the reference level of the center channel was set to three different reproduction levels. Note that in the third sub-experiment of experiment 3, the center channel was highpass filtered prior to calibration. The loudness of the centre channel was thus constant and equal for all signals in all three experiments. To achieve this calibration a Brüel & Kjær (B&K) 4134 pressure microphone was placed, upward facing, at the centre of the listening position at ear height (1.05m), without the listener's presence. This was connected via a B&K 2609 microphone amplifier to a B&K 2144 analyser with the 7638 loudness module.

The differences in sensitivity between the employed loudspeakers were aligned such that all loudspeakers had the same free-field sensitivity. This was performed in accordance with IEC 60268-5 [14] in the following manner. Loudspeakers were measured on-axis in an anechoic chamber at a distance of 1m employing a band limited pink noise signal with -3 dB corner frequencies at 200 Hz and 2000 Hz and 24 dB/octave roll-off.

## 1.2. The subjective task

A single subject participated in each session and was asked to adjust the level of an individual channel to be subjectively equal to that of the centre channel. A method of adjustment paradigm [15] was employed where the subject was free to switch between the centre channel and the test channel. The set-up allowed the subject to make the adjustments in fixed steps in the range 0.1-0.35 dB. The initial level of the channel was randomly set in the range of +2 to +6 dB or -2 to -6 dB compared to the center channel. This ensured that the initial level difference between the centre and variable channel was clearly perceivable. The subject was instructed to be facing forward during the entire session. The final gain, i.e. the subjective  $L_{diff\_S}$  value is recorded and represents the gain required to align the subjective level of the test channel to be equal to that of the reference channel. A block diagram of the subjective method of adjustment process is given in figure 6a.

Listeners were put through a three-stage procedure for the tests, consisting of

- a familiarisation and training session
- a training experiment
- the main experiments

At the initial stage, listeners were provided with oral and written instructions. They were presented with all the signals under consideration, and allowed to consider the task in hand. The use of the switching system was illustrated and listeners were allowed to test the system.

The training experiment was intended to have two functions: 1) train the listener for the task and further familiarise them with the test system and procedure, 2) test for listener reliability. This experiment consisted of a subset of the main experiment with all test signals employed with only the right hand front and surround channels. This data is not employed in the analysis presented in this paper.

## 1.3. Experiment 1: Source placement

The aim of experiment 1, reported in [2], was to study how listeners subjectively align levels under ideal multichannel reproduction conditions, as specified in ITU BS.775-1 [16]. Under these idealised conditions a symmetrical multichannel setup was placed into a symmetrical room, comprising identical loudspeakers placed at  $0^\circ$ ,  $\pm 30^\circ$ ,  $\pm 110^\circ$ . The actual test configuration is illustrated in figure 2. Whilst the setup is symmetrical within the room, the listeners head has direction properties in the horizontal plane, which can be associated with directional loudness properties. This has been researched by [5, 17, 18]. This factor is also investigated in this experiment. The identical experiment was setup at two sites, namely, NRC and B&O, where six listeners from the trained listening panels were employed. At the analysis stage one of the listeners at NRC was dropped from the analysis due to extreme error variance.

### 1.4. Experiment 2: Source distance and room influence

Experiment 2, reported in [2], attempts to examine the less ideal situation, as presented in figure 1(c). The actual test configuration is illustrated in figure 3. The study considers how both the asymmetries associated with the loudspeaker setup and the room can affect the level alignment. The net result is that certain loudspeakers will be strongly influenced by the room, e.g. the left speaker should encounter a corner loading, whilst others are little affected by the boundaries, i.e. the right surround channel. The source distances have also be exaggerated to emphasize source distance effect.

### 1.5. Experiment 3: Source directivity and bandwidth

The third set of experiments, reported in [3], aims to study the influence of loudspeaker directivity with a geometrically symmetrical setup, symmetrically placed in a room. Loudspeakers with four different directivity characteristics were employed, as presented in table 1. The effect of highpass filtering of the audio were also investigated (-3 dB @ 150 Hz, 6 dB/octave). A total of three sub-experiment were performed. The last of these also considered the effects of highpass filtering the reference center channel.

### 1.6. Experiment 4: Reproduction level and step size

This experiment, reported in [4], focuses upon the effects of the absolute reproduction level as a factor. The experiment was performed at three different calibration loudness levels of 15, 20 and 25 sones (Zwicker diffuse field assumption). Additionally, the minimum step size employed in the method of adjustment procedure was reduced to 0.1 dB, compared to previous experiments (0.2-0.35 dB).

## 2. OBJECTIVE METRICS

In order to facilitate the analysis of the subjective data, a set of measurements was performed that would allow for the detailed study of objective metric ( $L_{diff,O}$ ) of the sound field. To ensure that these measurements were as generic as possible, impulse responses (IR) were collected for each of the reproduction channels at each of the sites. The aim of this work has been to try to establish some correlation between objective and subjective data.

### 2.1. Impulse response measurements

Impulse responses were measured for each test system employing the MLSSA test system (<http://www.mlssa.com>) with an upward facing B&K 4134 microphone and associated conditioning amplifier. Nine points were measured in the horizontal plane, as illustrated in figure 7. The microphone spacing was chosen as 18 cm as this associates well with the average separation between the ears. In addition, the central microphone position was also employed, as during calibration. At each measurement point 16 measurements are averaged and stored at a sampling frequency of 95238 Hz, with a length of 688 ms. Having collected IR's, it is possible to calculate a broad range of objective responses by convolving the IR with the original test signals. This method was considered more convenient and flexible than making all measurements. Additionally, to provide a true estimate of the response at the ear a single set of head and torso simulator (B&K 4128) IR measurements was made for each reproduction channel at the listener position. The latter dataset is not considered in this paper.

The following metrics were considered of interest:

- Linear sound pressure level (dB), [19],
- A-weighted sound pressure level (dB), [19],

- B-weighted sound pressure level (dB), [19],
- C-weighted sound pressure level (dB), [19],
- D-weighted sound pressure level (dB), [20],
- Zwicker loudness, diffuse field assumption (sones), [13],
- Zwicker loudness, free field assumption (sones), [13],
- Moore loudness, diffuse field assumption (sones), [21],
- Moore loudness, free field assumption (sones), [21].

Two options present themselves as to how objective metrics might be calculated from the IR's and test signals and subsequently statically analysed. The first approach would be to directly calculate the objective metrics from the data. The second approach considers estimating an objective value  $L_{diff,O}$  that corresponds to aligning the test channel to an equal level as the reference center channel for any given metric.

When considering the means for statistical analysis, we are seeking a relationship between the subjective and objective domains. Our subjective data can be considered to reside within a “dB gain domain”. The raw metric data resides in various domains included the gain domain (e.g. SPL associated metrics) and the loudness domain (e.g. Zwicker and Moore loudness metrics). As the objective data lie within two domains, it might be difficult to estimate a linear association with the subjective data. The second approach to estimating the objective data would allow for all data to be transformed in the gain domain allowing for direct linear correlation (i.e. Pearson correlation) or linear regression analysis of all metrics. In this respect, it becomes possible to compare all subjective and objective data in a simplified manner. The second method was adopted.

Lastly, it was considered of interest to further study the importance of the measurement bandwidth on the correlation between the subjective and objective domains. This is a topic that was suggested by Bech [9]. To study this the objective metrics have been estimated in 6 different bandwidths. The raw data (all\_freq) contains frequencies above 50 Hz. Highpass filtering was applied at 250 Hz, 500 Hz, 1 kHz, 2 kHz and 3 kHz for which the analysis was performed in case.

## 2.2. Estimation of objective metrics

The overall process of objective metric estimation is presented in figure 6(b) and discussed here. To estimate the magnitude response for each channel and each signal, the IR's must first be combined with the samples. Each IR was interpolated<sup>1</sup> to 96 kHz. A time-frequency conversion was performed with an FFT hamming window of 48000 samples. Phase information for this task was not required. The test signals were then multiplied with the transfer function in the frequency domain to obtain the channel magnitude response for each signal and averaged over 5 seconds using one second blocks. This magnitude response is comparable to that which would have been directly measured in the room.

Estimation of the overall calibration gain,  $G_{cal}$ , for the measurement setup at each site was performed with the knowledge that the center channel was aligned to 20 soness<sup>2</sup> (Zwicker diffuse field assumption) with the center microphone for each test signal. This gain factors also contains gains associated with the samples and the reproduction system as well as the microphone sensitivity and

<sup>1</sup>the Matlab `interpft` function was employed.

<sup>2</sup>note that in experiment 4 the center channel was also calibrated to 15, 20 or 25 soness.

amplifier gain. This gain value is constant for all reproduction channels, as loudspeakers were calibrated for equal sensitivity, as described in section 1.1. Each of the models for the metrics itemized in section 2.1 have been implemented in Matlab. The Zwicker loudness models was used to calculate the overall calibration gain for the center channel with each signal by use of an optimization routine<sup>3</sup>. The routine was designed to converge when the loudness level was  $20 \pm 0.01$  sones.

To estimate the objective  $L_{diff\_O}$  values for each signal and each channel, the 1/3 octave spectra were calculated for each microphone position and averaged, with equal weighting, in the frequency domain. A 192-point amplitude spectrum was also estimated for input to the Moore loudness model [21].

The objective  $L_{diff\_O}$  value for each signal and channel are then estimated as illustrated in figure 6b. In words, the objective  $L_{diff\_O}$  value is the gain required to achieve the same reproduction level according to each metric, for each channel and signal. Once again an optimisation routine was employed to estimate these gain values.

### 3. ANALYSIS

The primary aim of the analysis is to ascertain the most suitable test signal and objective metric combination to predict the collected subjective data. For this purpose we have two possible tools. Firstly, we can perform a correlation analysis between the subjective and objective data to establish the signal/metric combination that provides the highest level of correlation. The second tool available to us is a linear regression analysis. As the objective metrics have been transformed into dB gain values, it now becomes possible to perform such a regression analysis with a known target function. The most suitable test signal/metric combination should results in regression line with a gradient of 1 and an intercept at 0 dB, explaining a large proportion of the variance in the data. Both of these analyses will now be presented.

#### 3.1. Correlation analysis

A Pearson correlation coefficient,  $r$ , is a suitable metric for this data, as both  $L_{diff\_O}$  and  $L_{diff\_S}$  are interval in nature. The correlation analysis was performed for all datasets between  $L_{diff\_O}$  and  $L_{diff\_S}$  in different bandwidths. The results for the 6 frequency bands are presented per signal in tables 3 and 4. Due to the mixed design of the subjective test, it should be noted that the sample size ranges from  $N = 227 - 264$ . Precise values are listed with the two-tailed significance levels for the correlation analysis in table 5.

Having performed the correlation analysis, we now need to establish whether or not there exist significant differences between any two  $r$ 's. For this purpose Fisher has developed a metric [22] which is discussed by Howell [23], and shown in equation 1. The highest  $r$  value is found to be 0.83 which is encouraging. Based upon this all values that are not significantly different have been estimated and are indicated by the grey shading in tables 3 and 4. Note that the significance of dependent on the sample size employed for each  $r$ .

$$z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}} \quad (1)$$

---

<sup>3</sup>the Matlab `fmin` routine was employed.

where

$$r' = (0.5) \log_e \left| \frac{1+r}{1-r} \right| \quad (2)$$

where  $r'$  is the transformed correlation coefficients under consideration according to equation 2.  $N$  is the total number of cases. A table of  $r'$  values can be found from ([22], pp.682).

### 3.2. Regression analysis

A linear regression analysis was performed for with the dependent variable  $L_{diff\_S}$  and independent variable  $L_{diff\_O}$  for each signal/metric combination in different bandwidths. The results of this study are presented in part in table 6, which illustrated the best fitting regression lines in each frequency band. The criterion defined for the “best fit” is a regression line with an intercept of  $a = 0$  dB and a slope of unity ( $b = 1$ ) as defined in equation 3. Regression line equations and associate 95% confidence intervals are also provided for both  $a$  and  $b$  in the table and the r-squared value, providing a measure of the explained variance in the regression line.

$$L_{diff\_S} = a + b * L_{diff\_O} \quad (3)$$

Figures 8 to 12 provide a summary the superior signal/metric combinations with scatter plots and fitted linear regression lines. Please note that the scales of the regression plots are not identical, in order to maximise presentation detail.

### 3.3. Regression analysis assumptions

Whilst a linear regression analysis can be performed with any two variables, we need to test whether it is fair to assume that a linear fitting is valid and meaningful. To do this we need to consider if our data meets the assumptions for the method and secondly how well a linear approximations fits the data .

The assumptions for linear regression are discussed in [23] and more extensively in [24] and will be considered. The two initial requirements for this method are homogeneity of variance in arrays and normality in arrays. In these respects our data are not perfect. In the most part the normality within arrays is quite respectable whilst the homogeneity of variance in arrays is not always fulfilled. Often we can see that there are only several discrete values of  $L_{diff\_O}$ , such that the conditional distribution assumption is not adhered to (see figures 8–12). Also we see that often there is a grouping of data around  $L_{diff\_O} = 0$  dB due to the experimental design which examined numerous symmetric configurations. Whilst this is not a problem in itself, it would have been more desirable to examine a more uniform distribution of  $L_{diff\_O}$  values.

To test whether it is reasonable to perform a linear regression on our data or whether higher order curve fitting is required, we can consider several metrics. Firstly, it is a good idea to look at the scatter plot of the data with the fitted line. The “big eyeball computer” will provide a good idea regarding the sense of the results. Some examples are illustrated in figures 13(a), 13(c) & 13(e) and 14(a), 14(c) & 14(e) which provide the linear, quadratic and cubic regression lines for a good and a poor case respectively. In figure 13 we can see that a linear fit is quite reasonable. The second metric to consider is the  $R^2$  value, which provides a measure of the amount of explained variance provided by the regression line. Lastly, the distribution of residuals provides a further indication of the goodness of fit, which is ideally normally distributed. Normal Q-Q plots illustrating the normality of distribution

of residuals are presented in figures 13(b), 13(d) & 13(f) and 14(b), 14(d) & 14(f) which provide the linear, quadratic and cubic regression lines for a good and a poor case respectively.

Considering these three metrics we can see that for the case illustrated in figure 13, the linear regression provides a good approximation of the data, with an  $R^2 = 0.627$ . The residuals are quite normally distributed. Higher order fitted curves only provide a slight improvement in the  $R^2$  values and marginally improve the normality of distribution of residuals. In the case illustrated in figure 14, we can observe that the linear regression does not describe the data well. This is supported by a low  $R^2$  value of 0.057, suggested that little of the data is explained by the regression line. Furthermore, considering figure 14(b) we can see that the data is not normally distributed and shows signs of significant non-linearity. Increasing the order of the regression to a cubic function does little to improve matters. In practice a  $\sim 5^{th}$  order curve fitting might be a benefit in this case, but this has not been studied. It is clear that the order of the curve fitting is an important issue here and its implications will be discussed further in the next section. Figure 14 exemplifies one of the least satisfactory solutions.

#### 4. RESULTS AND DISCUSSION

From the correlation analysis several things can be observed. Firstly, we see that for the all\_freq case (i.e. with a 50 Hz highpass filter frequency), we find that the  $r$ 's are very low for all signals and all metrics, in the order of 0.13-0.46. An example of the all\_freq highpass filter case is shown in figure 8, clearly illustrating a lack of linear fit. In general, significantly higher  $r$ 's are seen with highpass filtering  $\geq 250$  Hz. If we look across all the data in table 3 and 4, it can be found that a peak in the correlation values occurs for most signal/metric cases with a highpass filtering of 500 Hz for all signals. High  $r$ 's can be found for many of the signals in the order of 0.76-0.83. The consistent occurrence of high  $r$ 's occurs for all metrics with signal 5, with peak values with 500 Hz highpass filtering. Based upon this analysis alone it is not possible to isolate specific signal/metric combination for the prediction of the subjective responses.

When we try to analyse the regression data, the task is far greater and only a summary of the data is possible here. Firstly, we can state that the linear regression is a suitable means of predicting  $L_{diff_S}$  from  $L_{diff_O}$  in all cases presented except for the all\_freq case. In this situation measurements include all data above 50 Hz as presented in figure 14 and it is seen that a high order curve fitting would be required for this data. Clearly, the lower frequency information in measurements complicate matters and provide detail in the metrics that subjects are not highly sensitive to in the level calibration task. With 250 Hz–3 kHz highpass filtering, we find that for all metric/signal combinations a linear regression is an appropriate approximation and provides near normally distributed residuals and reasonable  $R^2$  values.

As discussed earlier, the transformation of the objective metric data into the dB gain domain allows us to seek the ideal signal/metric combination consisting of unity slope and intercept of zero. However, this is not the only metric of interest. Perhaps of greatest importance is the  $R^2$  value, describing the amount of variance explained by the regression line. Table 6 provides a summary of the “best fit” regression lines based upon the  $R^2$  values. The top metrics for each highpass filter frequency band are listed. Based on this we find that the greater variance is accounted for in the 500 Hz band, with peak values of 0.68–0.69, supporting the view provided by correlation analysis. However, in these cases the slope values are in the order  $b = 0.72$ . This is worse than in the best case, found in the 3 kHz band with  $b = 0.88$ . However, in practice this slope can simply be corrected for. These results support Bech's earlier view [9] that bandwidth is an important factor in level alignment. However, the frequency range proposed here is slightly lower than previously considered.

Considering the  $R^2$  value for all signal/metric combinations, we find that one signal provides  $R^2$

values that are consistently 4–15% higher than all other signals in all frequency bands from 250 Hz. This is signal 5, i.e. the constant specific loudness signal defined according to the Zwicker free-field loudness model.

In the 500 Hz band, we find that all B- and C-weighted SPL metric<sup>4</sup> provide the highest  $R^2 = 0.68 - 0.69$  and the loudness metric follow close behind with  $R^2$  values in the range 0.67. These latter metrics also have high slopes of 0.72 compared to the SPL metrics having slopes of 0.60. The results for this bandwidth are presented in figures 9-12. It can be noted when browsing through these figures that the superiority of any one single signal/metric combination is not so clear cut. Many combinations provide fair prediction of the subjective data. However, those discussed do appear to provide superior performance, even if only by a small margin.

These findings are in line with the results of work performed by Aarts [6, 7] and extend the concept to a wide range of practical usage situations.

## 5. CONCLUSIONS

A large number of data from subjective level alignment experiments, with noise signals in the range of 15–25 sones, of multichannel sound systems have been collected. A wide range of pragmatic domestic usage situations have been studied. Subjective and objective data have been correlated to attempt to find ideal test signal/metric combinations. The findings of this study have been that for a wide range of loudspeaker directivities, sensitivities, distances and positions, the following signal/metric combination provide superior prediction of subjective level calibration

- Constant specific loudness signal according the the Zwicker free field model
- Moore or Zwicker (diffuse or free field) loudness or B- or C-weighted SPL metrics
- Highpass filtering of the test signal from 500 Hz.

It is further concluded that with such level calibration methods, it possible the reasonably compensate for difference associated with different source distances and directivities. However, these findings need to be confirmed with further subjective experiments.

Lastly, it is found that the exclusion of low frequency information (in the range tested: 50–500 Hz) appears to improve the correlation between subjective level alignments and objective metrics. The exclusion of data below  $\sim 500$  Hz improves the objective prediction of the subjective level alignment in most cases considered.

## 6. FUTURE DIRECTIONS

Clearly there is a need for further study in this area. It would be of interest to study the regions of calibration that have not been studied in this research to establish the extend of validity of the proposed approaches. Subjective verification of alignment methods would also needed with true programme material, as opposed to noise test signals. Lastly the development of an automated calibration procedure based upon this data is of interest.

## 7. ACKNOWLEDGEMENTS

This work forms part of the studies of the Eureka 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications) project<sup>5</sup>. All members of the Medusa project are thanked for their

<sup>4</sup>in practice above 500 Hz both B- and C-weighted SPL metrics are identical in nature and thus should provide identical results.

<sup>5</sup>The Medusa project is a 3.5 years joint research project with the following partners: British Broadcasting Corporation, The Music Department of the University of Surrey, Nokia Research Centre, Genelec Oy, and Bang & Olufsen A/S.

comments and discussion throughout the project so far. Pekka Suokuisma is thanked for his efforts in developing the method and estimation of the objective metrics. Jan Abildgaard Pedersen is thanked for additional signal processing. The work was finalised during a Foreign Research Fellowship at the University of Aizu under the supervision of Bill Martens. The author would like to thank Bill for this useful comments on regression analysis and associated statistics. Tekes (Technology Development Center of Finland), the Danish DTI and the Nokia Foundation is thanked for funding support.

## REFERENCES

- [1] P. Suokuisma, N. Zacharov, and S. Bech, "Multichannel level alignment, part I: signals and methods," in *Proceedings of the Audio Engineering Society 105<sup>th</sup> Int. Conv.*, Audio Eng. Soc., 1998.
- [2] N. Zacharov, S. Bech, and P. Suokuisma, "Multichannel level alignment, part 2: The influence of signals and loudspeaker placement," in *Proceedings of the Audio Engineering Society 105<sup>th</sup> Int. Conv.*, Audio Eng. Soc., 1998.
- [3] S. Bech and N. Zacharov, "Multichannel level alignment, part 3: The effects of loudspeaker directivity and reproduction bandwidth," in *Proceedings of the Audio Engineering Society 106<sup>th</sup> Int. Conv.*, Audio Eng. Soc., 1999.
- [4] S. Bech and N. Cutmore, "Multichannel level alignment, part V: The effects of reproduction level, reproduction room, step size, and symmetry.," in *Proceedings of the Audio Engineering Society 108<sup>th</sup> Int. Conv.*, p. 24, Audio Eng. Soc., 2000.
- [5] N. Zacharov, "An overview of multichannel level alignment," in *Proceedings of the Audio Engineering Society 15<sup>th</sup> Int. Conf.*, pp. 174–186, Audio Eng. Soc., 1998.
- [6] R. M. Aarts, "Calculation of the loudness of loudspeakers during listening tests," *Journal of the Audio Engineering Society*, vol. 39, pp. 27–28, 1991.
- [7] R. M. Aarts, "Comparison of some loudness measures for loudspeaker listening tests," *Journal of the Audio Engineering Society*, vol. 40, pp. 142–146, 1992.
- [8] R. M. Aarts, *On the Design and Psychophysical Assessment of Loudspeaker Systems*. PhD thesis, Delft Technical University, 1995.
- [9] S. Bech, "Calibration of relative level differences of a domestic multichannel sound reproduction system," *Journal of the Audio Engineering Society*, vol. 46, pp. 304–313, 1998.
- [10] ITU-R, *Recommendation BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. International Telecommunications Union Radiocommunication Assembly, 1997.
- [11] R. Walker, "A controlled reflection listening room for multichannel audio," in *Proceedings of the 104<sup>th</sup> Convention of the Audio Engineering Society, Preprint 4645*, Audio Eng. Soc., May 16-19 1998.
- [12] IEC, *60268-1, Sound system equipment - Part 13: Listening tests on loudspeakers*. International Electrotechnical Commission, 1998.

- [13] ISO, 532, *Acoustics - Method for calculating loudness levels*. International Organization for Standards, 1975.
- [14] IEC, 60268-5, *Sound system equipment - Part 5: Loudspeakers Second Edition*. International Electrotechnical Commission, 1996.
- [15] B. L. Cardozo, "Adjusting the method of adjustment: SD vs DL," *Journal of the Acoustical Society of America*, vol. 37, no. 5, 1965.
- [16] ITU-R, *Recommendation BS.775-1, Multichannel stereophonic sound systems with and without accompanying picture*. International Telecommunications Union Radiocommunication Assembly, 1994.
- [17] D. W. Robinson and L. S. Whittle, "The loudness of directional sound fields," *Acustica*, vol. 10, pp. 74–80, 1960.
- [18] M. F. Sørensen, M. Lydolf, P. C. Frandsen, and H. Møller, "Directional dependence of loudness cues and binaural summation," *Proceedings of the 15<sup>th</sup> International Congress on Acoustics*, vol. 3, pp. 293–296, 1995.
- [19] IEC, 60651, *Sound level meters*. International Electrotechnical Commission, 1979.
- [20] IEC, 60537, *Frequency weighting for measurement of aircraft noise (D-Weighting)*. International Electrotechnical Commission, 1976.
- [21] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "A revision of Zwicker's loudness model," *Acta Acustica*, vol. 82, pp. 335–345, 1996.
- [22] R. A. Fisher, "On the probable error of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, pp. 3–32, 1921.
- [23] D. C. Howell, *Statistical methods for psychology*. Duxbury press, fourth ed., 1997.
- [24] D. C. Montgomery and E. A. Peck, *Introduction to linear regression analysis*. Series in probability and mathematical statistics, New York: John Wiley & Sons, second ed., 1992.
- [25] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, Heidelberg, Germany, 1990.

**FIGURES**

<b>Factor</b>	<b>Description</b>	<b>No. of levels</b>	<b>Degrees of freedom</b>
1. Signals	As describes in table 2	9	8
2. Loudspeakers	<ul style="list-style-type: none"> <li>• Genelec 1030A (reference)</li> <li>• B&amp;O Beolab 6000</li> <li>• Quad ESL63 (0°)</li> <li>• Quad ESL63 (90°)</li> </ul>	4	3
3. Listening rooms	<ul style="list-style-type: none"> <li>• NRC (ITU-R BS.1116-1)</li> <li>• B&amp;O (IEC 60268-13)</li> <li>• BBC (ITU-R BS.1116-1)</li> </ul>	3	2
4. Channels	5 channels (center reference)	4	3
5. Bandwidth	<ul style="list-style-type: none"> <li>• Full bandwidth</li> <li>• HP -3 dB @ 150 Hz, 6 dB/oct.</li> </ul>	2	1
6. Center ch. BW	<ul style="list-style-type: none"> <li>• Full bandwidth</li> <li>• HP -3 dB @ 150Hz, 6 dB/oct.</li> </ul>	2	1
7. Setup	<ul style="list-style-type: none"> <li>• Symmetrical</li> <li>• Assymetrical</li> </ul>	2	1
8. Level	<ul style="list-style-type: none"> <li>• 15 sones (Zwicker DF)</li> <li>• 20 sones</li> <li>• 25 sones</li> </ul>	3	2
9. Listeners	<ul style="list-style-type: none"> <li>• NRC - 6 trained listeners</li> <li>• B&amp;O - 6 trained listeners</li> <li>• BBC - 6 trained listeners</li> </ul>	3 x 6 (Nested)	3 x 5 (Nested)

Table 1: A summary of all the experimental factors and degrees of freedom

Signal name	High pass filter characteristics (Hz, dB/Oct.)	Low pass filter characteristics (Hz, dB/Oct.)	Comments
1.	700, 12	700,6	Commercially available signal
2.	250, 6	500, 6	A signal
3.	500, 18	2k, 18	Commercially available signal
4.			Zwicker constant specific loudness according to ISO 532 (diffuse field) [13]
5.			Zwicker constant specific loudness according to ISO 532 (free field) [13]
6.			Constant specific loudness according to Moore (diffuse field) [21]
7.			Uniform excitation noise according to Zwicker [25]
8.			Pink noise
9.			B-weighted pink noise

Table 2: A summary of the test signals employed in the subjective experiments [1]

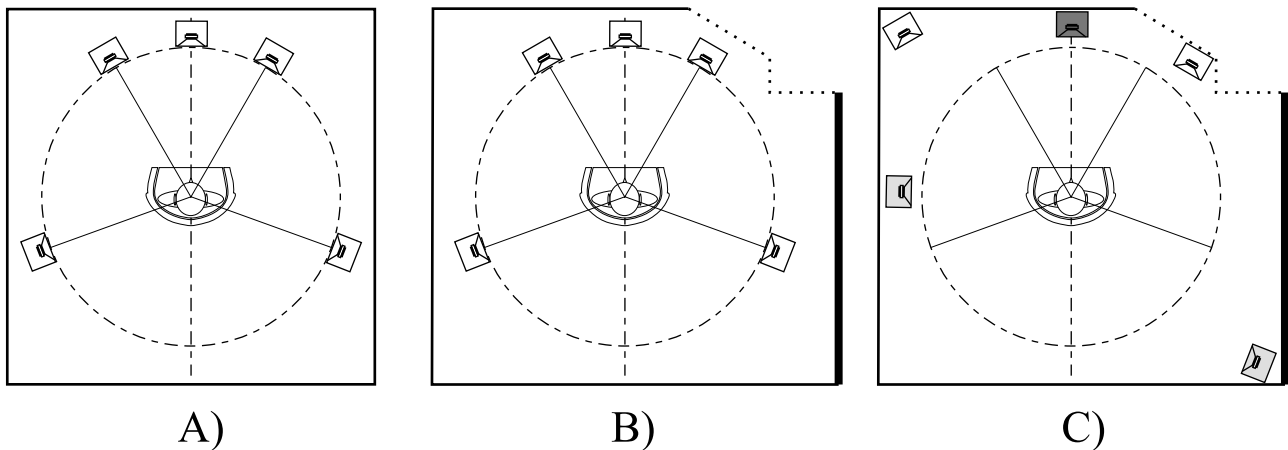


Figure 1: Some multichannel setups

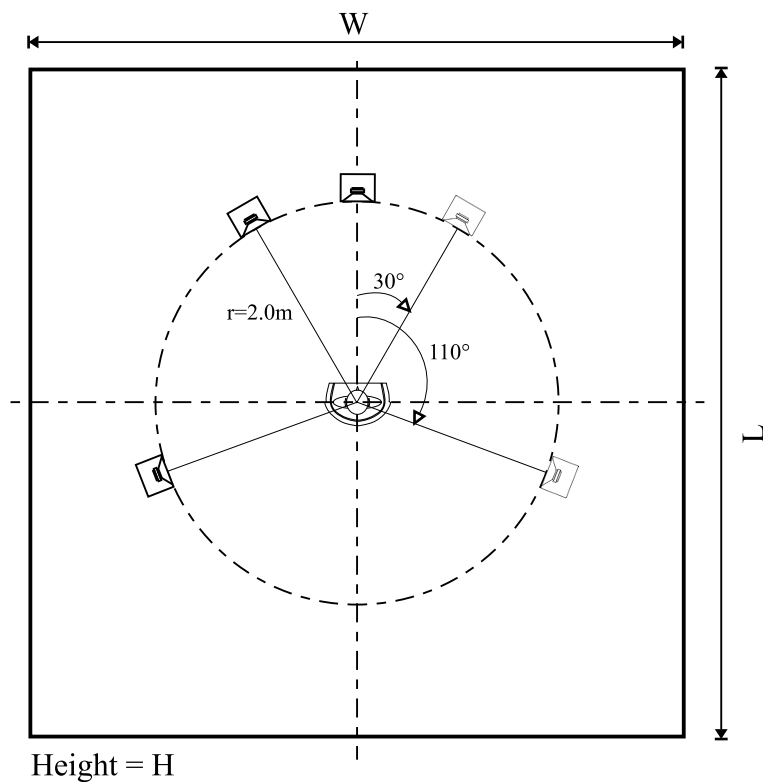


Figure 2: NRC &amp; B&amp;O listening room set-ups (experiment 1 &amp; 3)

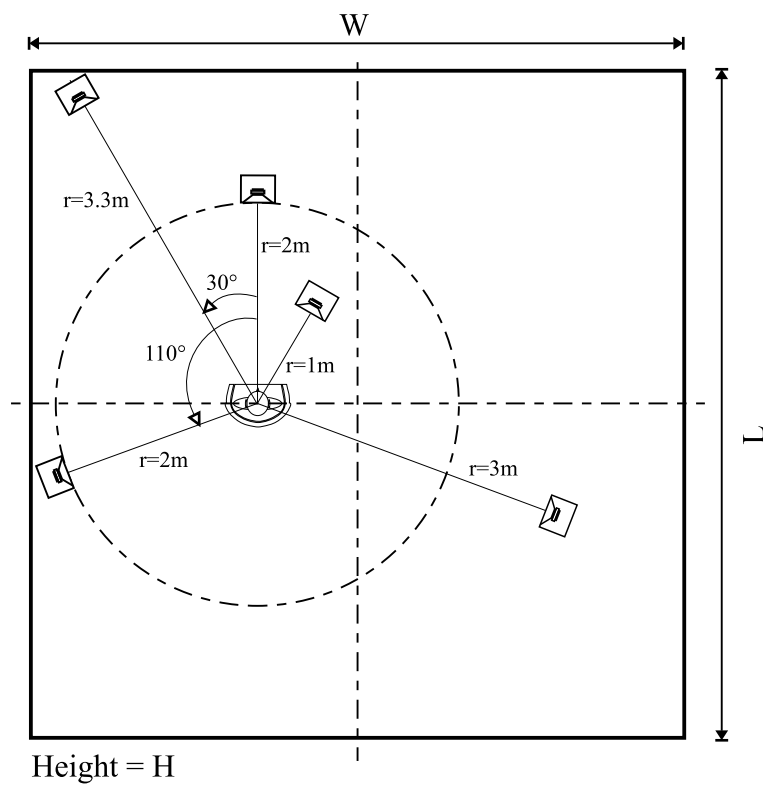


Figure 3: NRC listening room set-up (experiment 2)

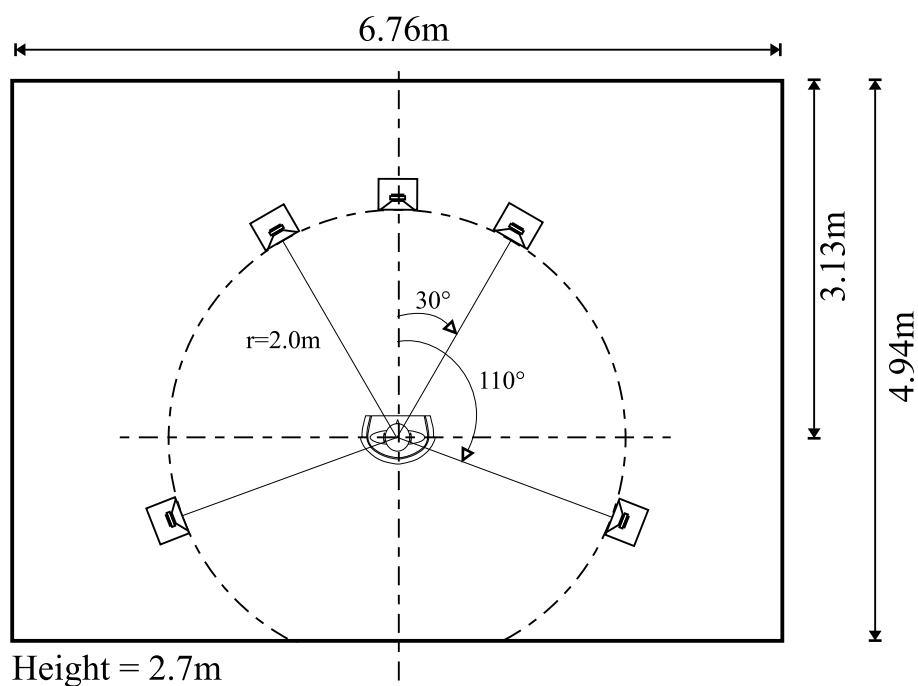


Figure 4: BBC listening room set-up (experiment 4)

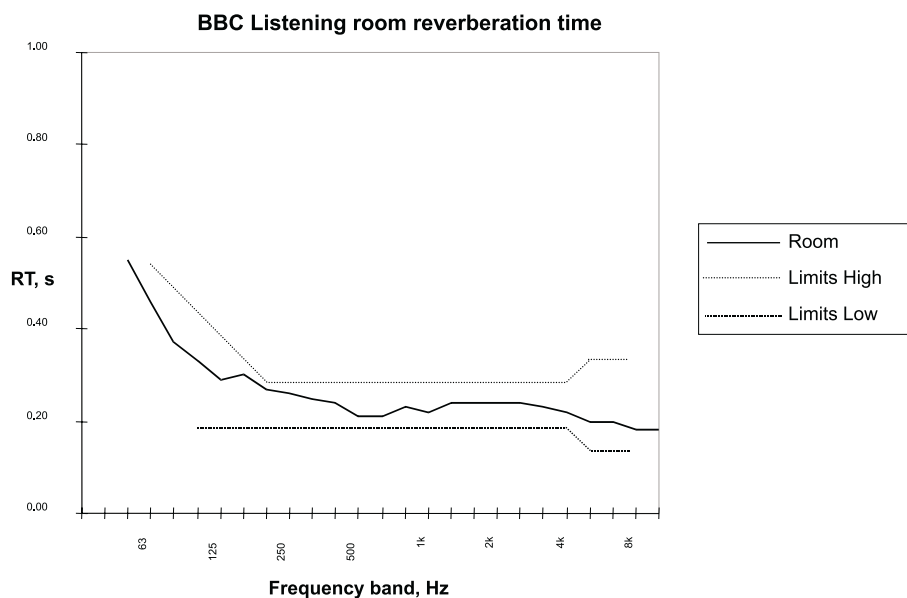


Figure 5: BBC listening room reverberation time characteristics (experiment 4). Limits defined by ITU-R BS.1116-1 [10]

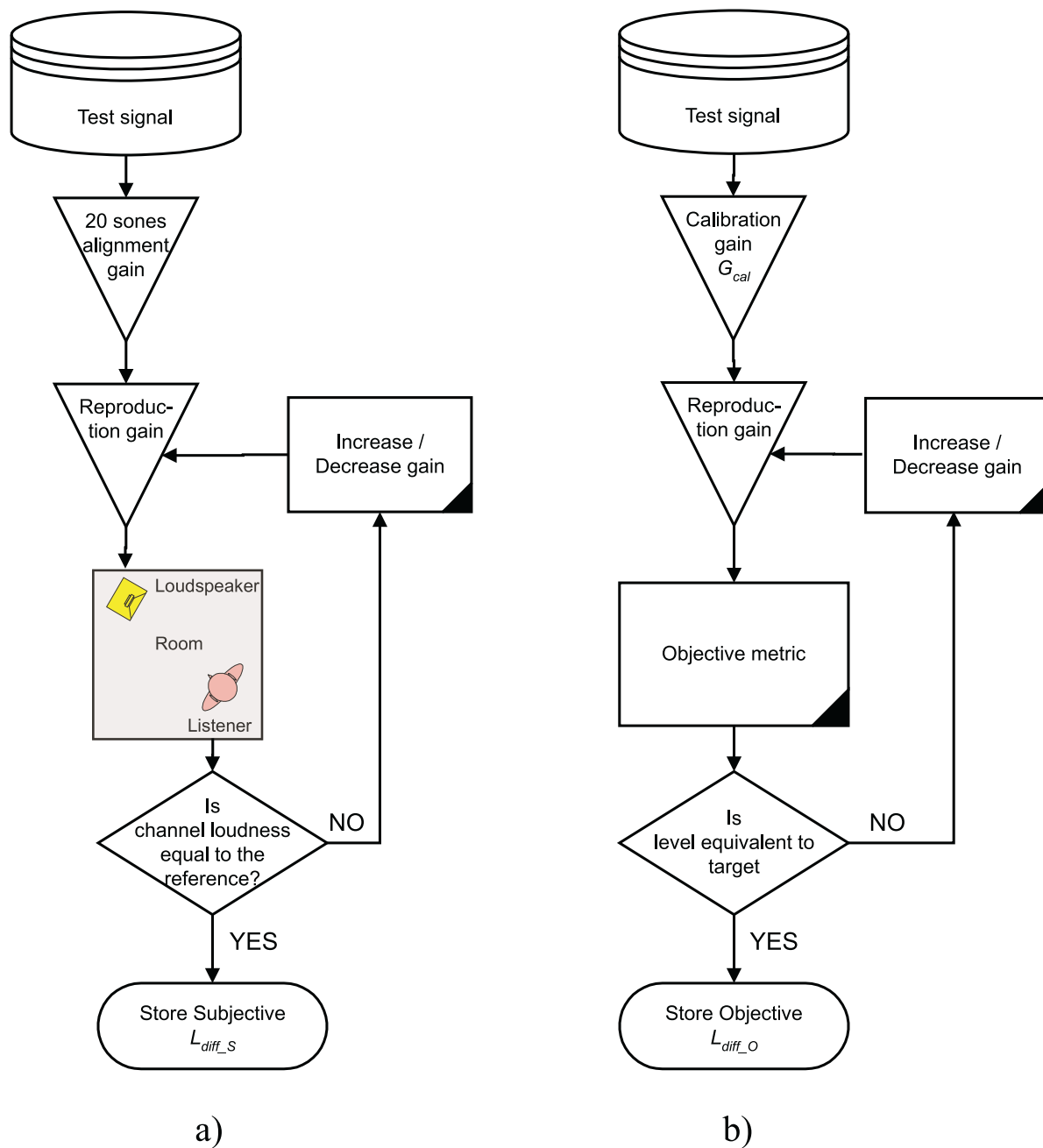


Figure 6: The block diagram of the level alignment process a) subjectively ( $L_{diff\_S}$ ) b) objectively ( $L_{diff\_O}$ ).

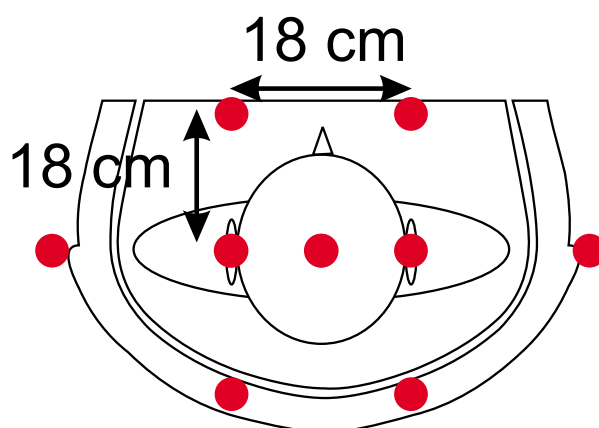


Figure 7: Objective measurement grid showing the positions of the measuring points.

Signal 1	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.23	0.72	0.74	0.73	0.73	0.71
ZWICK_FF	0.22	0.72	0.74	0.73	0.73	0.71
SPL_LIN	0.24	0.72	0.75	0.73	0.72	0.67
SPL_A	0.23	0.73	0.75	0.73	0.72	0.67
SPL_B	0.24	0.72	0.74	0.73	0.72	0.66
SPL_C	0.24	0.72	0.74	0.73	0.72	0.66
SPL_D	0.20	0.74	0.75	0.73	0.71	0.66
MOORE_DF	0.23	0.72	0.74	0.72	0.73	0.70
MOORE_FF	0.23	0.72	0.74	0.72	0.73	0.70

Signal 2	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.17	0.78	0.80	0.79	0.77	0.75
ZWICK_FF	0.16	0.78	0.79	0.78	0.77	0.75
SPL_LIN	0.13	0.74	0.77	0.75	0.68	0.65
SPL_A	0.19	0.78	0.78	0.75	0.68	0.65
SPL_B	0.16	0.75	0.77	0.75	0.68	0.64
SPL_C	0.13	0.74	0.77	0.75	0.68	0.64
SPL_D	0.15	0.77	0.77	0.73	0.67	0.64
MOORE_DF	0.17	0.79	0.79	0.78	0.76	0.74
MOORE_FF	0.16	0.79	0.79	0.78	0.76	0.73

Signal 3	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.28	0.75	0.76	0.76	0.76	0.72
ZWICK_FF	0.27	0.75	0.76	0.76	0.76	0.72
SPL_LIN	0.29	0.74	0.76	0.74	0.72	0.61
SPL_A	0.27	0.75	0.76	0.75	0.72	0.62
SPL_B	0.29	0.75	0.76	0.74	0.72	0.61
SPL_C	0.29	0.74	0.76	0.74	0.72	0.61
SPL_D	0.25	0.76	0.76	0.74	0.71	0.62
MOORE_DF	0.27	0.75	0.76	0.75	0.75	0.71
MOORE_FF	0.27	0.75	0.76	0.75	0.75	0.71

Signal 4	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.22	0.76	0.78	0.77	0.76	0.75
ZWICK_FF	0.20	0.76	0.78	0.77	0.76	0.74
SPL_LIN	0.28	0.72	0.74	0.71	0.68	0.64
SPL_A	0.25	0.74	0.75	0.72	0.70	0.66
SPL_B	0.23	0.73	0.76	0.72	0.70	0.66
SPL_C	0.19	0.73	0.76	0.72	0.70	0.66
SPL_D	0.23	0.73	0.74	0.72	0.70	0.67
MOORE_DF	0.20	0.76	0.78	0.76	0.76	0.74
MOORE_FF	0.19	0.76	0.78	0.76	0.75	0.73

Signal 5	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.19	<b>0.81</b>	<b>0.82</b>	<b>0.81</b>	<b>0.80</b>	0.78
ZWICK_FF	0.18	<b>0.81</b>	<b>0.82</b>	<b>0.81</b>	0.79	0.78
SPL_LIN	0.36	<b>0.81</b>	<b>0.81</b>	0.79	0.77	0.76
SPL_A	0.46	<b>0.82</b>	<b>0.82</b>	0.80	0.78	0.76
SPL_B	0.35	<b>0.82</b>	<b>0.83</b>	<b>0.81</b>	0.78	0.76
SPL_C	0.18	<b>0.82</b>	<b>0.83</b>	0.80	0.78	0.76
SPL_D	0.44	<b>0.82</b>	<b>0.82</b>	0.80	0.78	0.76
MOORE_DF	0.19	<b>0.81</b>	<b>0.82</b>	<b>0.81</b>	0.79	0.77
MOORE_FF	0.18	<b>0.81</b>	<b>0.82</b>	0.80	0.79	0.76

Table 3: Pearson correlation coefficients,  $r$ , for  $L_{diff-S}$  and  $L_{diff-O}$  as a function of signal (1–5) and objective metric for all data sets. Shaded cells indicate  $r$ 's that are not significantly different at a 95% confidence level compared to  $r = 0.83$ . Bold-italic cells indicate  $r \geq 0.81$ .

Signal 6	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.20	0.78	0.79	0.78	0.78	0.77
ZWICK_FF	0.18	0.78	0.79	0.78	0.78	0.77
SPL_LIN	0.35	0.73	0.74	0.73	0.71	0.70
SPL_A	0.41	0.75	0.76	0.75	0.73	0.72
SPL_B	0.34	0.75	0.76	0.75	0.73	0.72
SPL_C	0.16	0.75	0.76	0.75	0.73	0.72
SPL_D	0.39	0.75	0.76	0.75	0.73	0.72
MOORE_DF	0.20	0.78	0.79	0.78	0.78	0.76
MOORE_FF	0.18	0.78	0.79	0.78	0.77	0.75

Signal 7	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.20	0.74	0.76	0.76	0.76	0.75
ZWICK_FF	0.19	0.74	0.76	0.75	0.76	0.75
SPL_LIN	0.21	0.74	0.77	0.76	0.76	0.75
SPL_A	0.20	0.75	0.77	0.76	0.75	0.74
SPL_B	0.19	0.74	0.77	0.76	0.75	0.74
SPL_C	0.17	0.74	0.77	0.76	0.75	0.74
SPL_D	0.18	0.76	0.76	0.75	0.75	0.74
MOORE_DF	0.21	0.75	0.76	0.75	0.76	0.75
MOORE_FF	0.20	0.74	0.76	0.75	0.75	0.74

Signal 8	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.20	0.76	0.78	0.77	0.77	0.76
ZWICK_FF	0.19	0.76	0.78	0.77	0.77	0.76
SPL_LIN	0.17	0.74	0.79	0.76	0.75	0.73
SPL_A	0.21	0.77	0.78	0.76	0.74	0.73
SPL_B	0.17	0.74	0.78	0.76	0.74	0.73
SPL_C	0.14	0.73	0.78	0.76	0.74	0.73
SPL_D	0.18	0.76	0.77	0.75	0.73	0.72
MOORE_DF	0.20	0.76	0.78	0.77	0.77	0.75
MOORE_FF	0.20	0.76	0.78	0.77	0.77	0.75

Signal 9	ALL Freqs	250 Hz	500 Hz	1kHz	2kHz	3kHz
ZWICK_DF	0.20	0.74	0.76	0.76	0.76	0.75
ZWICK_FF	0.19	0.74	0.76	0.76	0.75	0.75
SPL_LIN	0.18	0.72	0.77	0.76	0.75	0.73
SPL_A	0.20	0.75	0.76	0.75	0.74	0.72
SPL_B	0.18	0.73	0.76	0.75	0.74	0.72
SPL_C	0.16	0.72	0.76	0.75	0.74	0.72
SPL_D	0.18	0.75	0.76	0.74	0.73	0.72
MOORE_DF	0.20	0.74	0.76	0.75	0.75	0.74
MOORE_FF	0.20	0.74	0.76	0.75	0.75	0.73

Table 4: Continuation of table 3. Pearson correlation coefficients,  $r$ , for  $L_{diff\_S}$  and  $L_{diff\_O}$  as a function of signal (6–9) and objective metric for all data sets. Shaded cells indicate  $r$ 's that are not significantly different at a 95% confidence level compared to  $r = 0.83$ . Bold-italic cells indicate  $r \geq 0.81$ .

	Sig. (2-tailed)	N
Signal 1	0.00	228
Signal 2	0.00	228
Signal 3	0.00	227
Signal 4	0.00	264
Signal 5	0.00	228
Signal 6	0.00	263
Signal 7	0.00	227
Signal 8	0.00	263
Signal 9	0.00	264

Table 5: Two-tailed significance levels and number of cases associated with correlation coefficients presented in tables 3 and 4.

Signal	Objective metric	HP filtering	Regression line: Ldiff_S = a + b * Ldiff_O	R-squared	95% confidence interval a (lower, upper)	95% confidence interval b (lower, upper)
Signal 5	SPL_A	All freq	0.19 + 0.47 * SPL_A	0.21	-0.017, 0.399	0.349, 0.589
Signal 5	SPL_D	All freq	0.17 + 0.45 * SPL_D	0.20	-0.035, 0.384	0.330, 0.567
Signal 6	SPL_A	All freq	-0.19 + 0.39 * SPL_A	0.16	-0.389, 0.008	0.284, 0.500
Signal 6	SPL_D	All freq	-0.18 + 0.38 * SPL_D	0.15	-0.380, 0.018	0.270, 0.485
Signal 5	SPL_A	250Hz	-0.26 + 0.56 * SPL_A	0.67	-0.397, -0.118	0.504, 0.606
Signal 5	SPL_B	250Hz	-0.32 + 0.59 * SPL_B	0.67	-0.461, -0.179	0.532, 0.639
Signal 5	SPL_C	250Hz	-0.34 + 0.59 * SPL_C	0.67	-0.481, -0.198	0.536, 0.643
Signal 5	SPL_D	250Hz	-0.25 + 0.55 * SPL_D	0.67	-0.388, -0.108	0.500, 0.602
Signal 5	Moore_DF	250Hz	-0.40 + 0.69 * moore_df	0.66	-0.549, -0.257	0.624, 0.753
Signal 5	Moore_FF	250Hz	-0.40 + 0.69 * moore_ff	0.66	-0.546, -0.253	0.629, 0.760
Signal 5	SPL_C	500Hz	-0.08 + 0.60 * SPL_C	0.69	-0.209, 0.058	0.545, 0.653
Signal 5	SPL_B	500Hz	-0.08 + 0.60 * SPL_B	0.68	-0.213, 0.054	0.542, 0.649
Signal 5	Moore_DF	500Hz	<b>-0.25 + 0.72 * moore_df</b>	<b>0.67</b>	<b>-0.391, -0.110</b>	<b>0.650, 0.782</b>
Signal 5	Moore_FF	500Hz	<b>-0.24 + 0.72 * moore_ff</b>	<b>0.67</b>	<b>-0.377, -0.096</b>	<b>0.656, 0.791</b>
Signal 5	Zwicker_FF	500Hz	-0.26 + 0.72 * zwick_ff	0.67	-0.397, -0.115	0.653, 0.786
Signal 5	Zwicker_DF	500Hz	-0.28 + 0.71 * zwick_df	0.67	-0.416, -0.135	0.646, 0.777
Signal 5	Zwicker_DF	1kHz	-0.21 + 0.74 * zwick_df	0.65	-0.356, -0.073	0.668, 0.807
Signal 5	Moore_DF	1kHz	-0.22 + 0.75 * moore_df	0.65	-0.363, -0.076	0.680, 0.825
Signal 5	Zwicker_FF	1kHz	-0.20 + 0.74 * zwick_ff	0.65	-0.345, -0.060	0.673, 0.815
Signal 5	SPL_B	1kHz	-0.03 + 0.58 * SPL_B	0.65	-0.521, 0.114	0.521, 0.632
Signal 5	Zwicker_DF	2kHz	-0.22 + 0.75 * zwick_df	0.64	-0.370, -0.077	0.677, 0.825
Signal 5	Moore_DF	2kHz	-0.21 + 0.81 * moore_df	0.63	-0.354, -0.058	0.725, 0.888
Signal 5	Zwicker_FF	2kHz	-0.20 + 0.76 * zwick_ff	0.63	-0.352, -0.058	0.682, 0.834
Signal 5	Moore_FF	2kHz	-0.18 + 0.81 * moore_ff	0.62	-0.333, -0.036	0.730, 0.897
Signal 5	Zwicker_DF	3kHz	-0.21 + 0.76 * zwick_df	0.61	-0.357, -0.056	0.685, 0.844
Signal 5	Moore_DF	3kHz	-0.18 + 0.86 * moore_df	0.60	-0.333, -0.027	0.767, 0.951
Signal 5	Zwicker_FF	3kHz	-0.18 + 0.78 * zwick_ff	0.60	-0.330, -0.026	0.694, 0.859
Signal 6	Zwicker_DF	3kHz	-0.57 + 0.72 * zwick_df	0.60	-0.714, -0.425	0.646, 0.791

Table 6: Summary of regression analysis. Top signal/metric combinations in each frequency band based upon a ranking of  $R^2$ . Shaded cells indicate the best fitting scenarios. Bold cells indicate high  $R^2$  and also high slope values.

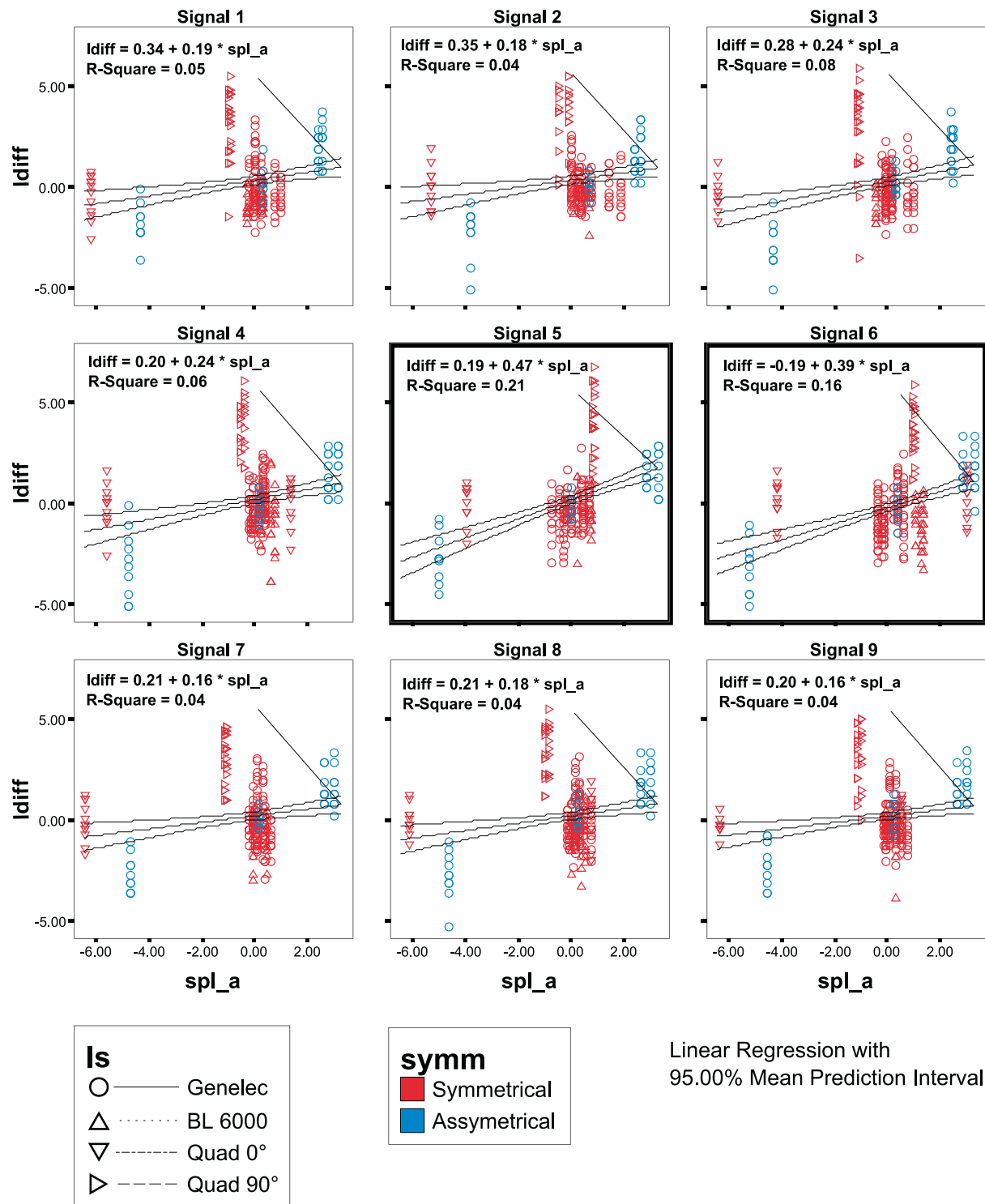


Figure 8: Scatter plots and linear regression lines for all (9) signals employing the C-weighted SPL metric with all\_freq data.

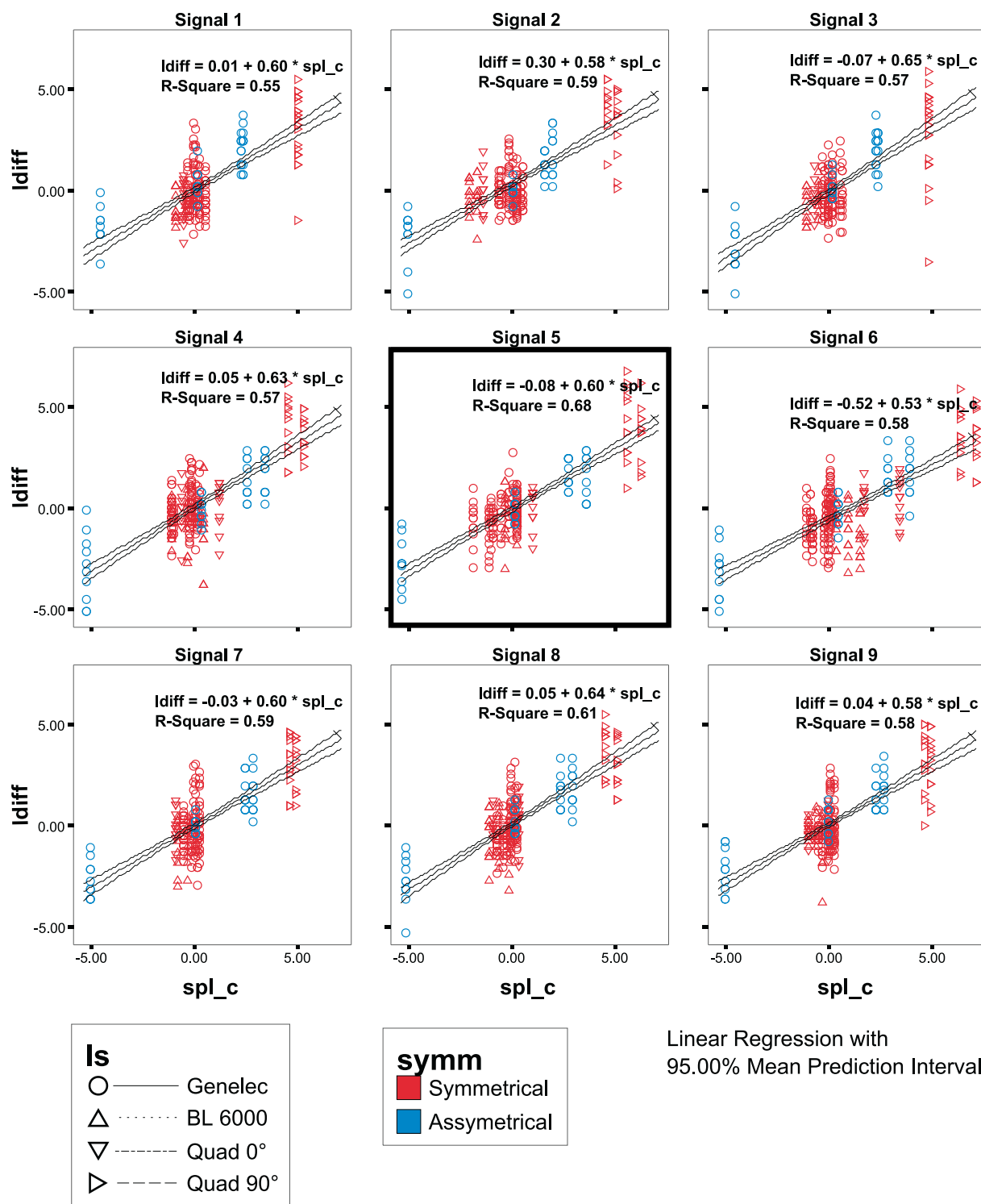


Figure 9: Scatter plots and linear regression lines for all (9) signals employing the C-weighted SPL metric with 500 Hz highpass filtering.

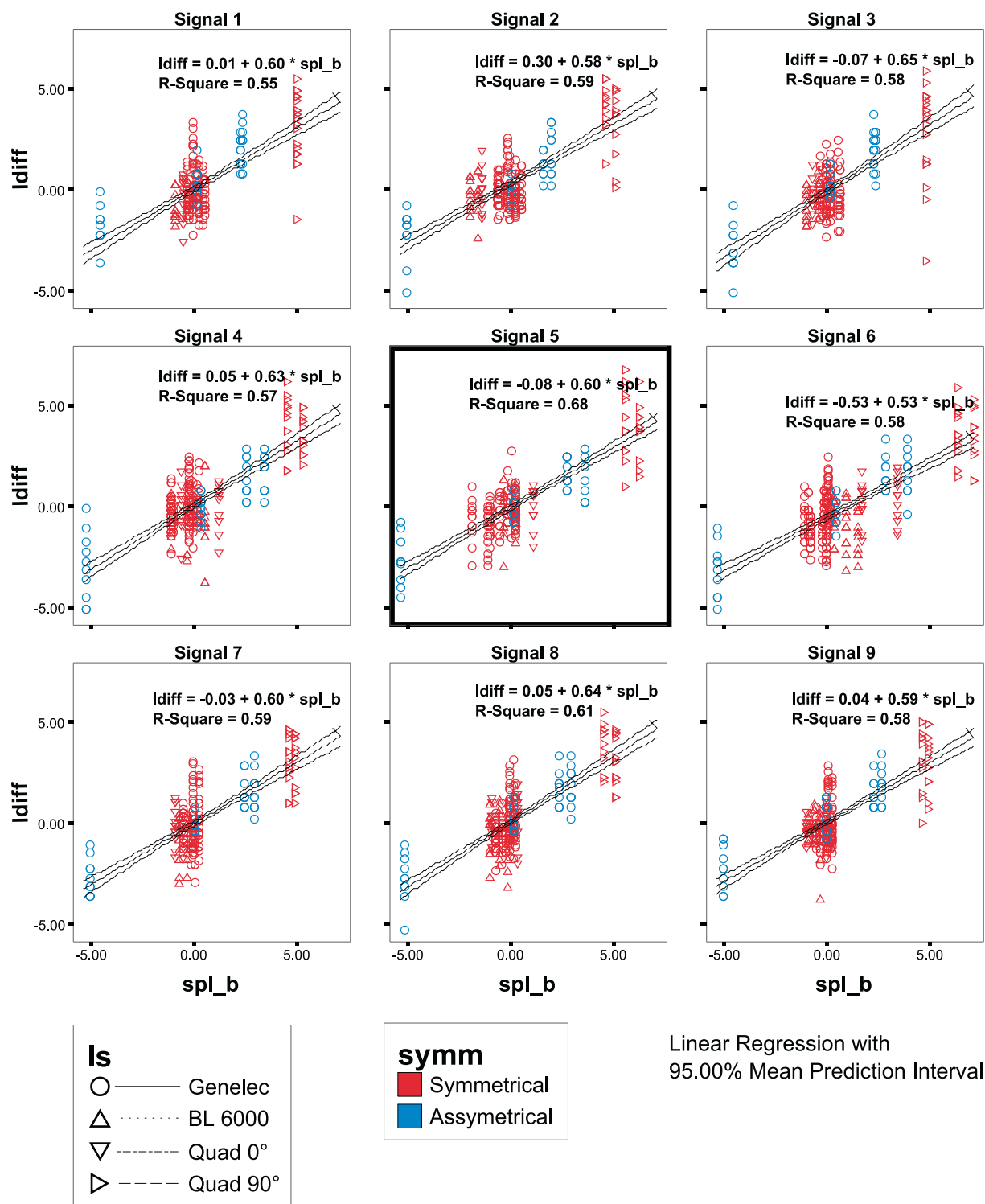


Figure 10: Scatter plots and linear regression lines for all (9) signals employing the B-weighted SPL metric with 500 Hz highpass filtering.

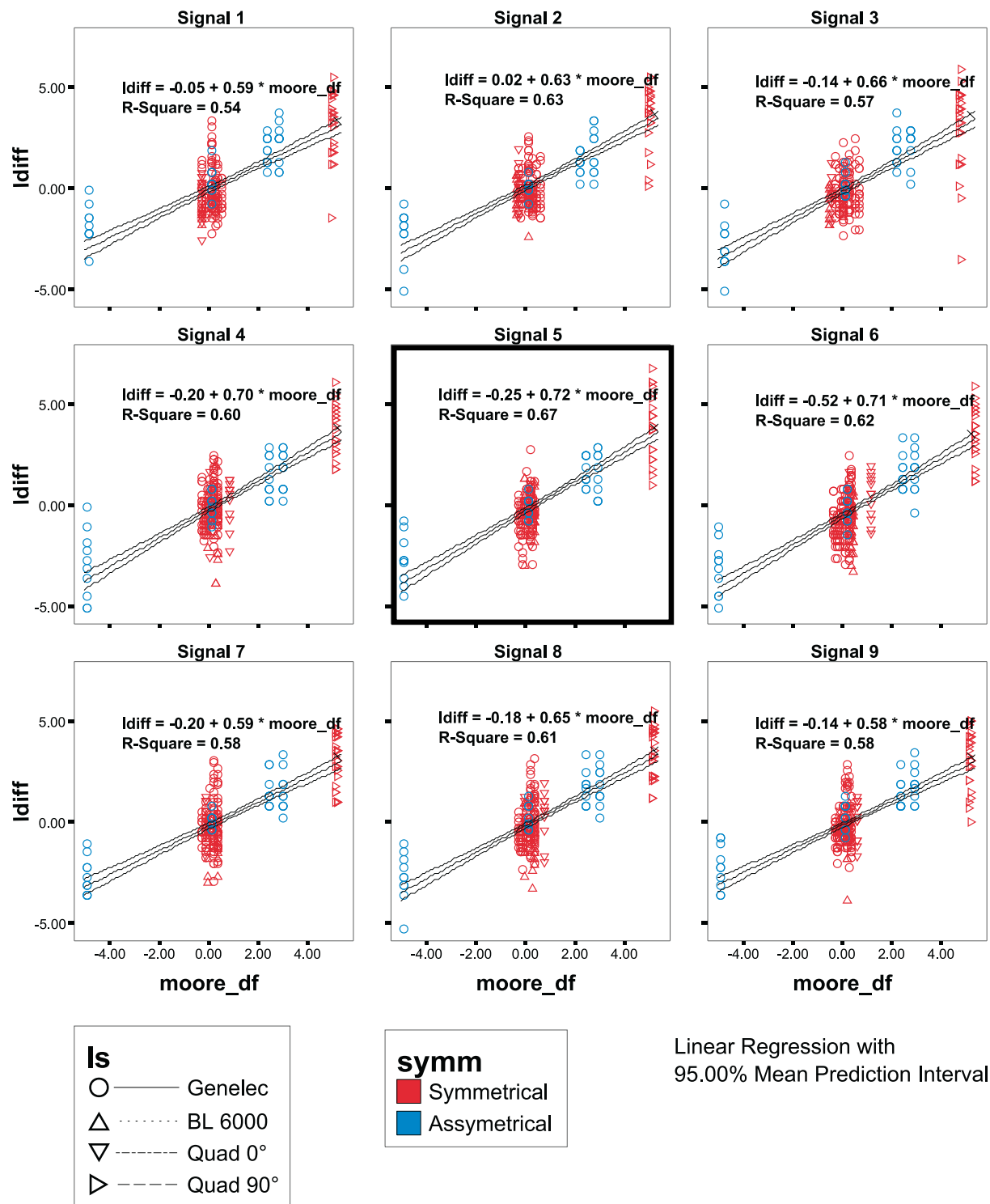


Figure 11: Scatter plots and linear regression lines for all (9) signals employing the Moore diffuse field metric with 500 Hz highpass filtering.

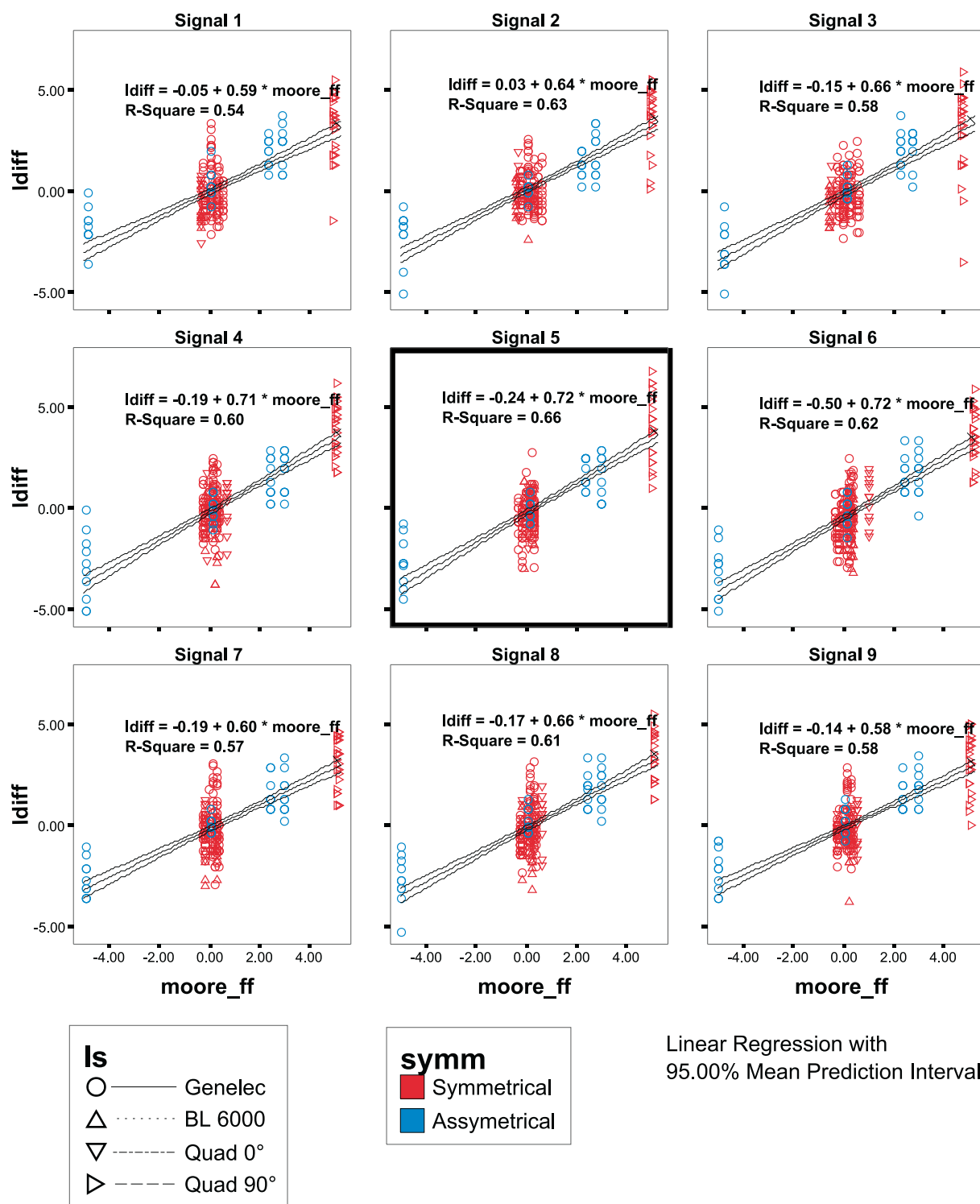
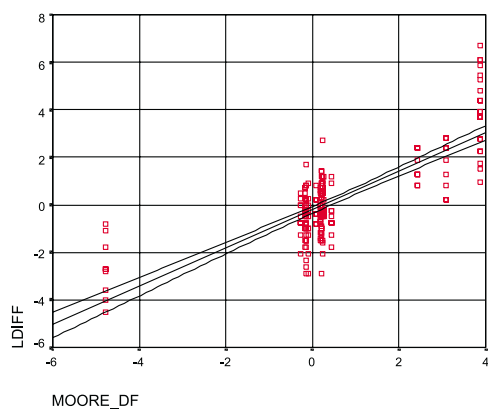
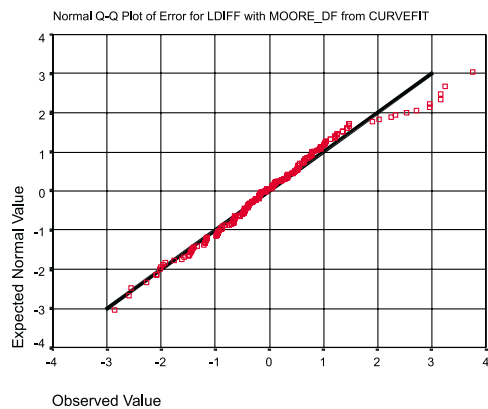
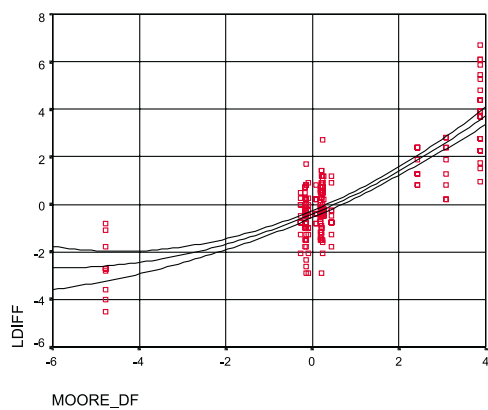
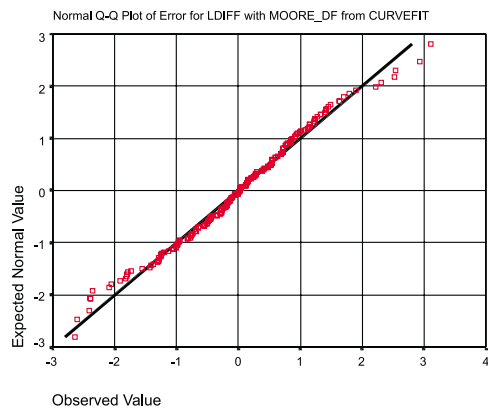


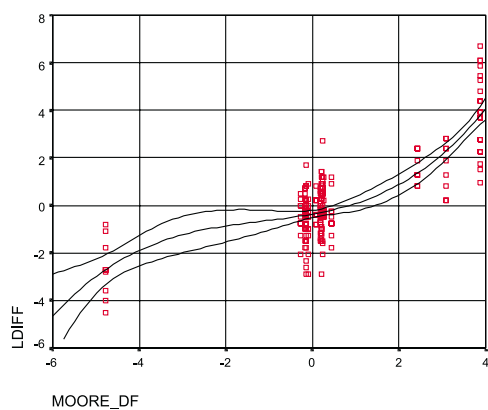
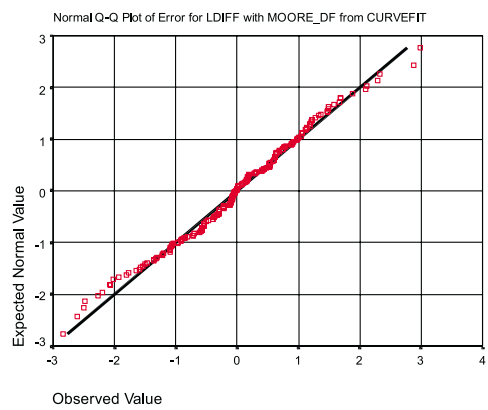
Figure 12: Scatter plots and linear regression lines for all (9) signals employing the Moore free field metric with 500 Hz highpass filtering.

(a) Linear regression line.  $R^2 = 0.627$ 

(b) Normal Q-Q plot of residuals of linear regression

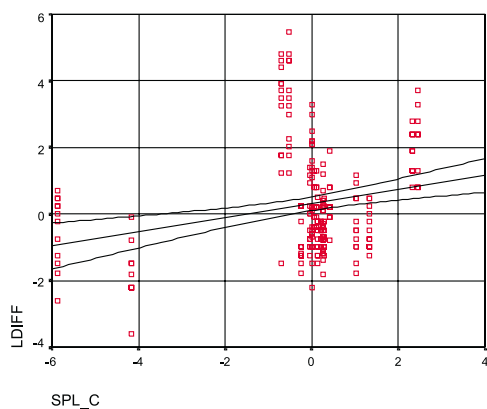
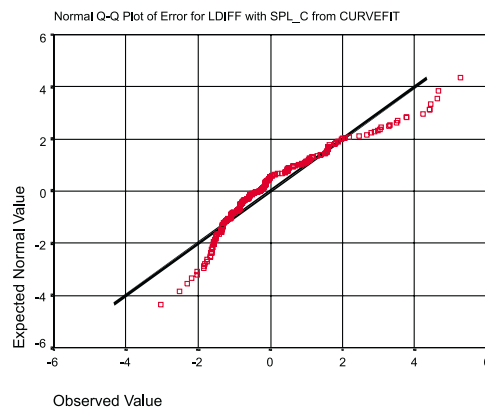
(c) Quadratic regression line.  $R^2 = 0.681$ 

(d) Normal Q-Q plot of residuals of quadratic regression

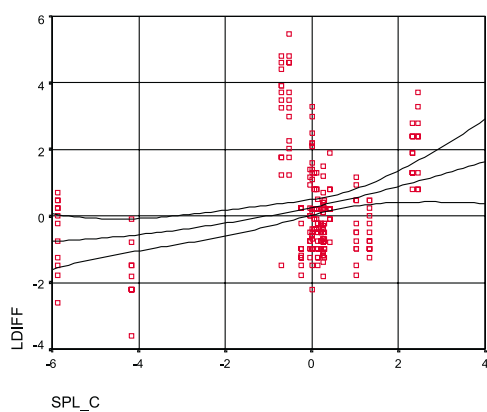
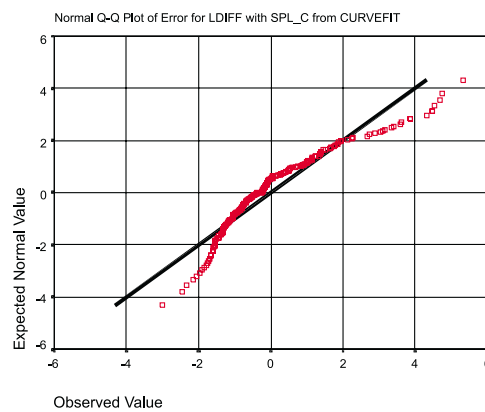
(e) Cubic regression line.  $R^2 = 0.690$ 

(f) Normal Q-Q plot of residuals of cubic regression

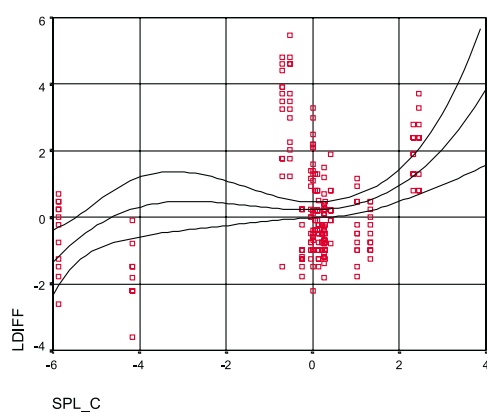
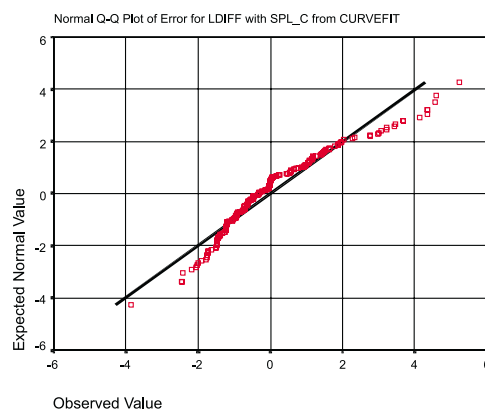
Figure 13: Higher order regression analyses for signal 5 and Moore\_DF metric in the 2kHz frequency band.

(a) Linear regression line.  $R^2 = 0.057$ 

(b) Normal Q-Q plot of residuals of linear regression

(c) Quadratic regression line.  $R^2 = 0.060$ 

(d) Normal Q-Q plot of residuals of quadratic regression

(e) Cubic regression line.  $R^2 = 0.081$ 

(f) Normal Q-Q plot of residuals of cubic regression

Figure 14: Higher order regression analyses for signal 1 and A-weighted SPL metric in the all\_freq band.