Spatialization Parameter Estimation in MDCT Domain for Stereo Audio

K. Suresh *sureshk@cet.ac.in*
Department of Electronics & Communication
Government Engineering College
Wayanad, Kerala, India, 670644

Akhil Raj R. *akhilraj.89@gmail.com*
Department of Electronics & Communication
College of Engineering, Thiruvananthapuram
Kerala, India, 695016

Abstract

For representing multi-channel audio at low bit rate parametric coding techniques are used in many audio coding standards. An MDCT domain parametric stereo coding algorithm which represents the stereo channels as the linear combination of the 'sum' channel derived from the stereo channels and a reverberated channel generated from the 'sum 'channel has been reported in literature. This model is inefficient in capturing the stereo image since only four parameters per sub-band is used as spatialization parameters. In this work we improve this MDCT domain parametric coder with an augmented parameter extraction scheme using an additional reverberated channel. We further modify the scheme by using orthogonalized de-correlated channels for analysis and synthesis of parametric stereo. A synthesis scheme with perceptually scaled parameter set is also introduced. Finally we present, subjective evaluation of the different parametric stereo schemes using MUSHRA test and the increased the perceptual audio quality of the synthesized signals are evident from these test results.

Keywords: Parametric Audio Coding, MDCT, Parametric Stereo.

1. INTRODUCTION

For multichannel audio compression with reasonable quality at low bit-rates, parametric coding has emerged as a suitable method with many potential applications [1,4,11]. In multichannel audio, significant amount of inter channel redundancies present along with perceptual irrelevancies and statistical redundancies. Effective removal of the inter channel redundancy and perceptual irrelevancy is required for low bit rate compression of multichannel audio. Considering the constituent channel data individually, we can apply mono audio compression methods to remove perceptual irrelevancies and statistical redundancies. To remove inter channel redundancies, cross channel prediction based methods were suggested [2]. However, in most of the cases, the correlation between the channels is low and the cross channel prediction based methods will not result in significant compression [3]. Another strategy for suppressing the inter channel redundancy is through parametric coding. In parametric coding, the encoded data consists of a mono 'sum' channel derived from the individual channels and model parameters representing the spatialization cues. Binaural cue coding (BCC) introduced in [8,9,10] and parametric stereo coding method introduced in [11] are examples for parametric multichannel audio coding. BCC uses inter channel level difference (ICLD), inter channel time difference (ICTD) and inter channel coherence (ICC) as parameters for spatial audio modeling. In the encoder, a 'sum' channel is derived by adding the individual channels followed by energy equalization. The 'sum' channel is compressed using any of the existing audio coding algorithms. The spatialization parameters are extracted in a frame by frame basis and quantized for compact representation. In the decoder, the multichannel audio is synthesized using the 'sum' channel

signal and the spatialization parameters; to synthesis the required level of correlated side channels, decorrelated signals are generated from the 'sum' channel. In the human auditory system, the processing of binaural cues is performed on a non-uniform frequency scale [12,13]. Hence, in order to estimate the spatial parameters from the given input signal, it is desirable to transform its time domain representations to a representation that resembles this non uniform frequency scale. This transformation is achieved either by using a hybrid Quadrature Mirror Filter (QMF) bank or by grouping a number of bands of a uniform transform such as DFT [9,11]. However, in practice, for audio coding purpose, spectral representations such as Modified Discrete Cosine Transform (MDCT) are used which has the advantage of time domain alias cancellation and better energy compaction. Hence additional filter bank analysis or transform is needed for the parameter extraction in the encoder and for the synthesis in the decoder. The spatialization parameter extraction and 'sum' channel formation is done as a pre-processing step in the encoder; conversely, the stereo synthesis is a post-processing step in the decoder. Similarly, the de-correlated channel generation in the decoder is done either by time domain convolution or equivalent DFT domain multiplication [9,11]. MDCT domain analysis and synthesis of reverberation for parametric coding of stereo audio has been proposed in [14]. Spatialization parameter extraction and stereo synthesis from the 'sum' channel are done in the MDCT domain. For parameter estimation, the MDCT coefficients are divided in to twenty two non-uniform blocks and an analysis by synthesis scheme in the MDCT domain is used. The stereo channels are approximated to the linear combination of the 'sum' channel and a reverberated channel derived from the 'sum' channel. Four parameters are extracted from each block and encoded as the side information. The spatialization parameters such as ICTD, ICLD and ICC are not estimated directly. Instead, the de-correlated channel used for stereo synthesis in the decoder is generated in the encoder and it is used to estimate the synthesis coefficients through least square approximation method. The parametric coder realized using this parameter extraction method is capable of achieving reasonably good quality stereo audio. In this paper, we propose an improved parametric extraction scheme in the MDCT domain using three different approaches. In the first scheme we use two reverberated channel instead of the single reverberated channel as proposed in [14]. This results in better modeling of spatialization cues which is reflected in the perceptual evaluation. In the second scheme, we use sub-band wise mutually orthogonalized sum channel and reverberated channels for parameter extraction and synthesis. In the third scheme, psychoacoustically weighted parameter extraction scheme is introduced. Performance evaluations of proposed methods are conducted through listening tests.

2. SPATIALIZATION PARAMETER EXTRACTION AND STEREO SYNTHESIS USING MULTIPLE REVERBERATED CHANNELS

Formation of 'sum' channel through down-mixing, generating reverberated channels from the 'sum' channel and parameter extraction are the functions of a parametric stereo encoder. We follow the methods as used in [14] for in our encoder as described below. The block diagram for the MDCT domain parametric stereo encoding section is shown in Figure 1. In the first step, MDCT

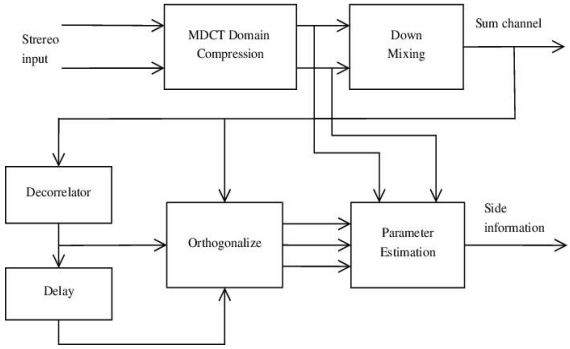

Figure 1: Block Diagram of MDCT Domain Parametric Stereo Encoder.

of the stereo channels $(x_1(n)$ and $x_2(n))$ are computed. The stereo signals in the MDCT domain are then down-mixed to form a 'sum' channel $S(k)$. A de-correlated channel $S_r(k)$, is derived from the sum channel. Additional de-correlated channel is generated by delaying the de-correlated channel. The frequency domain signals $S(k), S_r(k)$ and a delayed version of the de-correlated channel $S_{rd}(k)$ are used to extract the spatialization parameters.

2.1 Down Mixing to Sum Signal

For the analysis of stereo signal 1024 point MDCT of the input stereo channels are computed. In order to imitate the features of human auditory system which is more sensitive to lower frequency bands than higher bands, the MDCT coefficients are grouped into bands with different spectral resolution. We use a partitioning method same as that suggested by C. Faller in [1] in which the MDCT coefficients are grouped into 22 non-overlapping frames of different lengths. The partition details are given in table 1. A sub-band by sub-band energy equalized 'sum' signal is obtained from stereo transforms by down mixing. The MDCT of the j^{th} sub-block of the energy equalized sum signal is given by Equation 1.

$$S^j(k) = c_j\{X_1^j(k) + X_2^j(k)\}, \quad for\ all\ k \tag{1}$$

$$c_j = \sqrt{\frac{\frac{1}{2}\sum_k ((X_1^j(k))^2 + (X_2^j(k))^2)}{\sum_k (X_1^j(k) + X_2^j(k))^2}} \tag{2}$$

$X_1^j(k)$ and $X_2^j(k)$ are the MDCT coefficients of j^{th} sub band. The energy equalization factor c_j given by equation 2 is introduced to make the energy of the 'sum' sub-band signal equal to the average energy of the corresponding stereo sub-bands.

Partition	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}
Boundary	0	8	16	24	32	48	64	80	96	128	160	192
Partition	B_{13}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}	B_{18}	B_{19}	B_{20}	B_{21}	B_{22}	
Boundary	224	256	288	320	384	448	512	576	640	768	1024	

TABLE 1: Partition boundaries of sub-bands in MDCT domain.

2.2 Synthesis of De-correlated Channels

For parametric coding in MDCT domain, the synthesis parameters are not derived directly from spatialization parameters like ICLD, ICTD and ICC. Instead, we estimate them using the sum signal, a de-correlated 'sum' signal and a delayed de-correlated signal. The same de-correlated signals are created in the decoder to synthesize the stereo signal. The de-correlated signal is computed using the method suggested by J. Breebaart et al., in [11], i.e., by convolving the 'sum' signal with an all pass de-correlation filter whose impulse response is given by equation 3.

$$h(n) = \sum_{m=0}^{\frac{N}{2}} \frac{2}{N} \cos\left(\frac{2\pi mn}{N} + \frac{2\pi m(m-1)}{N}\right) \tag{3}$$

De-correlated signal is generated by the convolution of 'sum' signal and $h(n)$. The convolution is implemented in the MDCT domain following the method described in [16]. The MDCT of the de-correlated 'sum 'is given by

$$S_r^{mdct}(k) = | H_{u,8N}^{dft}(2k+1) | \{S^{mdct}(k)\cos(\theta_{2k+1}) + S^{mdct}(k)\sin(\theta_{2k+1})\}, \quad 0 \le k \le N-1 \quad (4)$$

where $H_{u,8N}^{dft}(2k+1)$ is the 8N point DFT of up-sampled $h(n)$. The length of the de-correlation filter is selected such that L << N for which the error due to aliasing is irrelevant [16]. We use de-correlation filter of length L=40 for N=1024.

2.3 Parameter Estimation in MDCT Domain

For the parameter extraction, we use 'sum' signal $S(k)$, de-correlated signal $S_r(k)$ and delayed de-correlated signal $S_{rd}(k)$ which is obtained by shifting $S_r(k)$ by a certain number of samples. The MDCT coefficients representing the synthesized stereo channels $(X_1^j(k), X_2^j(k))$ are given by the linear combination

$$X_i^j(k) = a_i^j S^j(k) + b_i^j S_r^j(k) + S_{rd}^j(k), \quad (5)$$

where *i=1,2* and *1 ≤ j ≤ 22.*

To estimate the synthesis parameters a_i^j , b_i^j and c_i^j the inner product between the transforms of stereo signals $S(k), S_r(k)$ and $S_{rd}(k)$ are computed.

$$\begin{aligned} d_i^j &= \langle X_i^j(k) S^j(k) \rangle \\ e_i^j &= \langle X_i^j(k) S_r^j(k) \rangle \\ f_i^j &= \langle X_i^j(k) S_{rd}^j(k) \rangle \end{aligned} \quad (6)$$

An approximation of the stereo signal is obtained by adding the projections of $X_1^j(k)$ and $X_2^j(k)$ on $S^j(k), S_r^j(k)$ and $S_{rd}^j(k)$ as

$$P_i^j(k) = d_i^j S^j(k) + e_i^j S_r^j(k) + f_i^j S_{rd}^j(k) \quad (7)$$

Further amplitude scaling of $P_i^j(k)$ is done to equalize the energy in the synthesized sub-bands to that of the original sub-band energy by multiplying with the scaling factors computed as below

$$X_i^j(k) = s_i^j P_i^j(k) \quad (8)$$

where the scale factors s_i^j i=1,2 are given by

$$s_i^j = \frac{\sum_k (X_i^j(k))^2}{\sum_k (P_i^j(k))^2} \quad (9)$$

The synthesis parameters are obtained by multiplying the inner products obtain in the first step with the corresponding scale factors.

$$a_i^j = s_i^j d_i^j \qquad b_i^j = s_i^j e_i^j \qquad c_i^j = s_i^j f_i^j \quad (10)$$

The synthesis parameters are compressed using uniform quantizer and encoded as side information.

2.4 Stereo Decoding

The block diagram for the decoder is shown in Figure 2. The MDCT coefficients of the synthesized stereo are reconstructed from the linear combination of the equalized 'sum' signal $S(k)$, and de-correlated signals $S_r(k)$ and $S_{rd}(k)$. The de-correlated and its delayed signals are obtained using the same procedure followed in the encoder. Side information forms the weights for the linear combination.

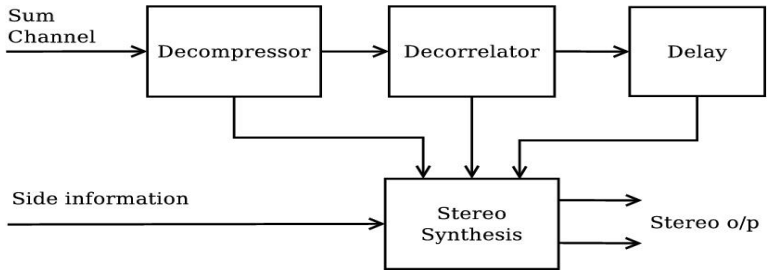

FIGURE 2: Block diagram of parametric stereo decoder.

2.5 Spatialization Parameter Estimation using Orthogonalized De-correlated Channels

De-correlated channels derived from the 'sum' channel through filtering are not perfectly correlated. To further reduce the correlation between them we performed sub-band wise orthogonalization of the channels in the MDCT domain. The 'sum' signal $S(k)$, de-correlated signal $S_r(k)$, and delayed de-correlated signal $S_{rd}(k)$, are modified through Gram-Schmidt orthogonalization procedure. Orthogonalization of these signals is performed in every sub-band to make

$$\begin{aligned} \langle S^j(k)\ S_r^j(k)\rangle &= 0 \\ \langle S^j(k)\ S_{rd}^j(k)\rangle &= 0 \\ \langle S_r^j(k)\ S_{rd}^j(k)\rangle &= 0 \end{aligned} \qquad (11)$$

where j = 1,2,..,22 denotes the sub-band indices $S^j(k), S_r^j(k)$ and $S_{rd}^j(k)$ the corresponding orthogonalized 'sum', de-correlated and delayed de-correlated signals. We used these signals for parameter extraction. The decoder is also modified to include the orthogonalization steps performed in the encoder.

2.6 Stereo Decoding with Orthogonalized De-correlated Channels

The block diagram of the stereo decoder is shown in Figure 3. The decoder receives $S^j(k)$ and spatialization parameters $a_i^j(k), b_i^j(k)$ and $c_i^j(k)$. From this, first we obtains the de-correlated signals $S_r^j(k)$ and $S_{rd}^j(k)$. Then sub-band wise orthogonalization is done and the decoder synthesizes back the stereo channels as a linear combination of these orthogonalized signals using the spatialization parameters.

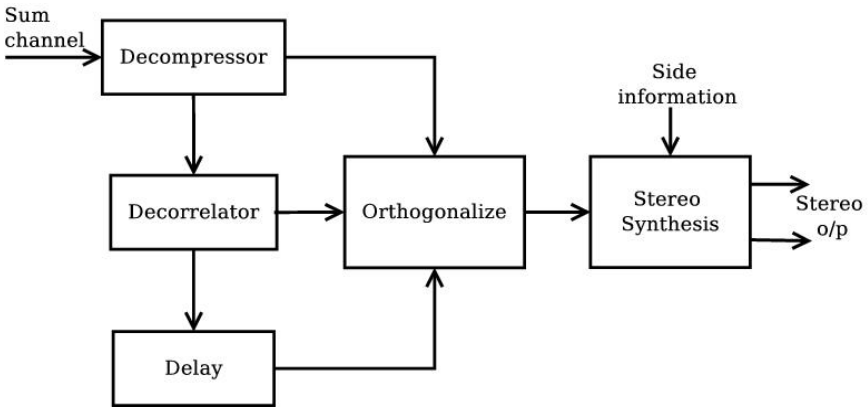

FIGURE 3: Stereo decoder for synthesizing stereo from orthogonalized 'sum' and reverberated channels.

2.7 Psychoacoustically Weighted Parameter Estimation Scheme

Masking threshold estimated through psychoacoustic analysis is used for determining perceptually irrelevant components of audio signals and extensively used in perceptual audio compression algorithms [17]. A higher masking threshold indicates higher noise masking capability of the frequency component. Components with lower masking threshold are more sensitive to quantization noise. We used masking threshold estimate to scale the spatialization parameters. Low weightage is given for spatialization parameters representing frequency components with higher masking thresholds since they are less sensitive components. Coefficients corresponding to lower masking threshold are amplified to give them more weightage. We used MDCT domain masking threshold estimation scheme presented in [18] for computing the masking threshold estimate $T_i(k)$ of the every audio frame. Scaling factors $m_i(k)$ for frequency components are obtained by uniform linear interpolation of $[\min(T_i(k))\ \max(T_i(k))]$ into the range [1.5 0.5] such that minimum weightage is given for frequency components having maximum masking threshold as given below.

$$N_i(k) = \left\lceil 1023\left(\frac{\max(T_i(k)) - T_i(k)}{\max(T_i(k)) - \min(T_i(k))} \right) \right\rceil \tag{12}$$

$$m_i(k) = 0.5 + \frac{N_i(k)}{1023} \tag{13}$$

This scaling factor $m_i(k)$ is used to scale the spatialization parameters in the stereo analysis stage as given in equation

$$a_i^j(k) = s_i^j(k)\, m_i(k) \langle X_i^j(k)\, S^j(k) \rangle$$

$$b_i^j(k) = s_i^j(k)\, m_i(k) \langle X_i^j(k)\, S_r^j(k) \rangle$$

$$c_i^j(k) = s_i^j(k)\, m_i(k) \langle X_i^j(k)\, S_{rd}^j(k) \rangle \tag{14}$$

where *j=1,2,..,22* denotes the sub-band indices, i the channel number, $s_i^j(k)$ the sub-band wise energy equalizing factor and $m_i(k)$ scaling function obtained from the masking threshold estimate. The method for stereo synthesis from parameters is same as in the case of stereo synthesis using multiple reverberated channels.

3. SUBJECTIVE EVALUATION RESULTS

To evaluate the perceptual quality of the encoded audio signal using the proposed algorithm, listening test was conducted. Six listeners participated in the test. The listeners are asked to evaluate both the spatial audio quality as well as other audible artifacts. In a MUSHRA test [19], the listeners had to rate the relative perceptual quality of ten processed items against original excerpts in a 100-point scale with 5 anchors. Tests are conducted with high quality headphone in a quiet room. The following items are included in the test.

- The original as the hidden reference.
- A low-pass filtered (cut off frequency of 7 kHz) mono channel derived from the original.
- Stereo audio compressed by MPEG-2 AAC with TNS and M/S stereo enabled at a rate of 64 kbps.
- Stereo signal synthesized using uncompressed 'sum 'signal and synthesis parameters estimated with single reverberated channel.
- Stereo signal synthesized using uncompressed 'sum 'signal and synthesis parameters estimated with two reverberated channel.

3.1 Perceptual evaluation of parametric stereo generated using multiple reverberated channels

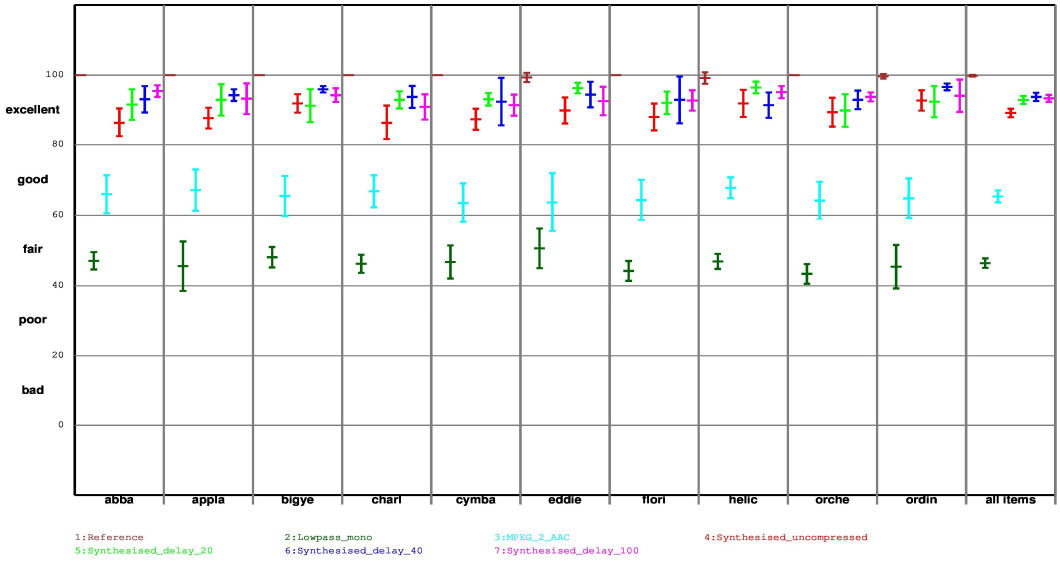

FIGURE 4: MUSHRA Scores for Average and 95% Confidence Intervals for stereo synthesis.

In the case of stereo synthesized using two reverberated channel, we have considered three different delay conditions (20 ms, 40 ms and 100 ms) for the second delay channel. Thus, there are seven clips including the hidden references for each channel in the subjective evaluation. All clips had a resolution of 16 bits per sample and sampling rate of 44.1 kHz. The list of stereo clips used for the subjective evaluation is shown in Table 2.

The average MUSHRA scores obtained for the seven version of each clip are shown in Table 3. The average score for all the clips combined is also given. The synthesized audio with two reverberated signals performed better than the synthesized signal with single reverberated signal. The average MUSHRA score for the synthesized signal with reverberated channels having no delay and 40 ms delay is 93.8 while that of single reverberated channel is 89.3. The performance of 20 ms delayed reverberated channel and 100 ms delayed channel are also very close to that of the 40 ms delayed case.

Sl. No.	Index	Name	Origin/Artist
1	abba	Abba	SQAM Database
2	appla	Applause	SQAM Database
3	bigye	Big Yellow	Counting Crows
4	charl	Charlies	Danny O'Keefe
5	cymba	Cymbal	Radiohead
6	eddie	Eddie Rabbit	SQAM Database
7	flori	Florida Sequence	Pre-echo test case
8	helic	Helicopter	bAdDuDeX
9	orche	Orchestra	Dave Mathews band
10	ordin	Ordinary World	Duran Duran

TABLE 2: List of excerpts used in the subjective listening test.

Name	Original	Low pass Mono	MPEG 2 AAC	Synth Uncomp-ressed	Synth. Delay 20	Synth. Delay 40	Synth. Delay 40
Abba	100	47	66	86.5	91.7	93.2	95.5
Applause	100	45.5	67.2	87.8	93	94.3	93.3
Big Yellow	100	48	65.5	92	91.3	96	94.3
Charlies	100	46.2	66.8	86.5	93	93.8	91
Cymbal	100	46.7	63.5	87.5	93.2	92.5	91.5
Eddie Rabbit	99.3	50.5	63.7	90	96.3	94.5	92.7
Florida Sequence	100	44.2	64.3	88.2	92.2	93	92.8
Helicopter	99.2	46.8	67.8	92	96.5	91.5	95.2
Orchestra	100	43.3	64.2	89.5	90	93	93.8
Ordinary World	99.7	45.3	64.8	92.8	92.5	96.7	94.2
All Items	**99.8**	**46.4**	**65.4**	**89.3**	**93**	**93.8**	**93.4**

TABLE 3: Average MUSHRA scores for test clips generated using multiple reverberated channels.

3.2 Perceptual evaluation of parametric stereo generated using orthogonalized signals

To evaluate the performance of stereo synthesized using orthogonalized de-correlated channels subjective evaluation was conducted with following audio clips.

- The original as the hidden reference.
- A low-pass filtered (cut off frequency of 7 kHz) mono channel derived from the original.
- Stereo audio compressed by MPEG-2 AAC at a rate of 32 kbps.
- Stereo signal synthesized using energy equalized 'sum' signal, de-correlated signal and its delayed version (with delay of 40 samples) weighted by synthesis parameters.
- Stereo signal synthesized using orthogonalized and energy equalized 'sum' signal, de-correlated signal and its delayed version (with delay of 40 samples) weighted by synthesis parameters.

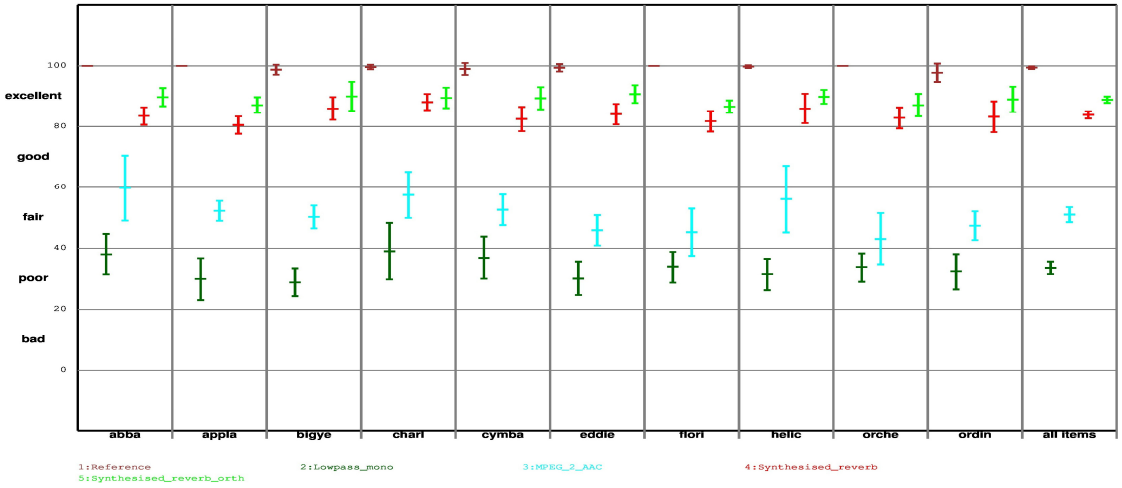

FIGURE 5: MUSHRA Scores for Average and 95% Confidence Intervals for audio clips.

Name	Original	Low pass mono	MPEG 2 AAC	Reverb. delay	Orthogonal Delay
Abba	100	38.9	59.8	83.5	89.8
Applause	100	29.9	52.2	80.5	87.1
Big Yellow	98.8	28.8	50.2	86	90
Charlies	99.6	39	57.5	88.1	89.5
Cymbal	99	36.9	52.6	82.5	89.4
Eddie Rabbit	99.4	30	45.9	84.1	90.8
Florida Sequence	100	33.8	45.2	81.8	86.6
Helicopter	99.8	31.4	56.1	86	89.9
Orchestra	100	33.6	43	82.9	87.1
Ordinary World	97.8	32.2	47.4	83.2	89
All Items	**99.4**	**33.4**	**51**	**83.9**	**89.9**

TABLE 4: Average MUSHRA scores of audio clips synthesized using orthogonalized signals.

Stereo signal synthesized using energy equalized 'sum' signal, de-correlated signal and its delayed version(with delay of 40 samples) weighted by synthesis parameters. Stereo signal synthesized using orthogonalized and energy equalized 'sum' signal, de-correlated signal and its delayed version(with delay of 40 samples) weighted by synthesis parameters. The average MUSHRA scores assigned by the listeners for these clips are shown in Table 4. The average score for all the clips combined is also given. MUSHRA scores for average and 95% confidence intervals are plotted in Figure 5.

Results clearly show that orthogonalizing the 'sum' and reverberated channels for spatialization parameter extraction and stereo synthesis have resulted in improving the perceptual quality of the synthesized audio. The average MUSHRA score for the synthesized stereo has increased from

83.9 to 89.9 when orthogonalized signals were used for parameter extraction and stereo synthesis.

3.3 Perceptual evaluation of parametric stereo generated using scaled parameters

Test audio clips were generated by scaling spatialization parameters extracted using estimated masking threshold values. This perceptual weighting of parameters is done for the stereo synthesis using multiple reverberated channels with and without orthogonalization. We used the following items for listening test:

- The original as the hidden reference.
- Stereo audio compressed by MPEG-2 AAC at a rate of 32 kbps.
- Stereo signal synthesized using energy equalized 'sum' signal, de-correlated signal and synthesis parameters.
- Stereo signal synthesized using energy equalized 'sum' signal, de-correlated signal and its delayed version (with delay of 40 samples) with scaled parameters.
- Stereo signal synthesized using orthogonalized and energy equalized 'sum' signal, de-correlated signal and its delayed version (with delay of 40 samples) with scaled parameters.

The average MUSHRA scores obtained for different test audio clips and average score obtained for all clips are shown in Table 5. MUSHRA scores for average and 95% confidence intervals are

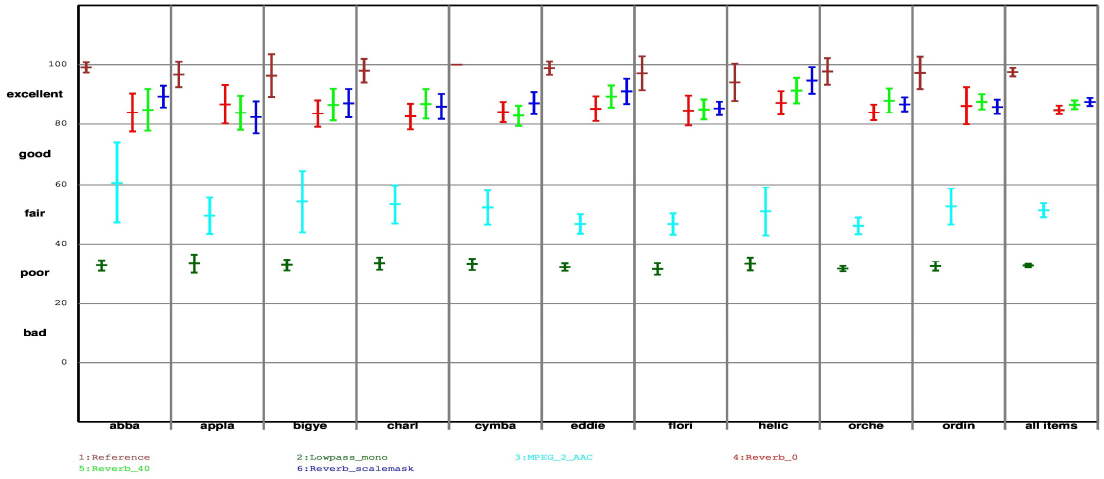

FIGURE 6: MUSHRA Scores for average and 95% confidence Intervals obtained for synthesized audio clips using scaled spatialization parameters.

plotted in Figure 6 The results of the listening test shows that the perceptual quality of the synthesized stereo using two reverberated channels was 86.7 whereas including masking threshold also to scale the spatialization parameters resulted in an average score of 87.6.

3.4 Discussion

It is not surprising that the quality of the encoder increases as the number of parameters is increased. Use of an additional delayed reverberated channel for stereo analysis and synthesis increases the perceptual quality at the cost of increased side information rate. The delayed signal helps the parametric model to capture the late reverberated part of the original signal. The reverberation filter length is 40 samples which is approximately equal to 1 ms and a delay of 40 samples effectively extends the filter length to 80 samples and that results in better modeling of the spatial cues. The number of parameters representing spatialization cues is increased to three per sub-band for each channel and this result in a better approximation for the stereo audio.

Name	Original	Low pass mono	MPEG 2 AAC	Synth. Uncompressed	Synth. Delay 40	Perceptually Weighted
Abba	99.1	32.8	60.6	84	84.9	89.4
Applause	96.8	33.4	49.5	86.8	83.9	82.4
Big Yellow	96.4	32.9	54.2	83.6	86.6	87.1
Charlies	98	33.4	53.4	82.6	86.9	86
Cymbal	100	33.1	52.2	84.1	82.9	87.1
Eddie Rabbit	98.9	32.1	46.8	85.2	89.4	91.1
Florida Sequence	97.1	31.5	46.8	84.6	85	85.4
Helicopter	94.1	33.2	51	87.2	91.4	94.8
Orchestra	97.8	31.6	46.1	84	88	86.8
Ordinary World	97.2	32.5	52.6	86.2	87.6	85.9
All Items	**97.5**	**32.6**	**51.3**	**84.8**	**86.7**	**87.6**

TABLE 5: Average MUSHRA scores of synthesized audio clips using scaled Spatialization parameters.

Perceptual tests shows better quality for stereo synthesized using additional reverberated channel delayed by 40 samples with an average MUSHRA score of 93.4 whereas that for stereo synthesized using single reverberated channel was 89.3. When we orthogonalize the de-correlated signals we expect better spatial modeling at the cost of additional computation. MUSHRA test reveals a better performance of stereo synthesized using orthogonalized signals with an average score of 89.9 against the average score of 83.3 obtained for stereo synthesized non orthogonalized reverberated channels. Stereo synthesis using perceptually scaled spatialization parameters produce marginal improvement in perceived quality at the cost of increased complexity. When evaluated simultaneously, the average MUSHRA score is 87.6 for scaled parameter synthesis while that of multiple de-correlated channels is 86.7.

4. CONCLUSION

Methods for parametric stereo coding in MDCT domain using sum and two de-correlated channels have been introduced. Three methods are used for estimating spatialization parameters. It can be seen that stereo synthesized using ‘sum’ and two de-correlated channels has a better perceptual quality than that synthesized using ‘sum’ and a single de-correlated channel. The quality of the encoder has increased when orthogonalized signals were used for parameter extraction and stereo synthesis. Orthogonalization makes the three signals used in parameter extraction and stereo synthesis independent of each other. But the computational complexity of encoder as well as decoder will be increased due to the additional orthogonalization process. In method 3, spatialization parameters were further modified using scaling functions obtained from masking threshold and results in marginal improvement in the perceptual quality of the synthesized audio.

5. REFERENCES

[1] C. Faller, “Parametric Coding of Spatial Audio,” *Swiss Federal Institute of Technology Lausanne (EPFL), PhD Thesis, No. 3062,* 2004.

[2] D. Yang, H. Ai, C. Kyriakakis, ans C.C. J. Kuo, “An inter channel redundancy removal approach for high quality multichannel audio compression,” in *AES convention,* Los Angeles, CA, Sept 2000.

[3] S. Kuo and J.D. Johnston, “A Study of Why Cross Channel Prediction is Not Applicable to Perceptual Audio Coding," IEEE Sig. Proc. Letters, vol. 8, No. 9, pp 245-247, Sep. 2001.

[4] J. Herre, et.al, “The reference Model Architecture for MPEG Spatial Audio Coding," in 118th AES convention, Barcelona, Spain May 2005, Preprint 6447.

[5] J.D. Johnston, and A.J. Ferreira, “Sum Difference Stereo Transform Coding,” in Proc. IEEE ICASSP-92, San Francisco, vol. 2, pp. 569-572, March 1992.

[6] Christian R. Helmrich, Pontus Carlsson, Sascha Disch, Bernd Edler, Johannes Hilpert, Matthias Neusinger, Heiko Purnhagen, Nikolaus Rettelbach, Julien Robilliard, and Lars Villemoes, “Efficient Transform Coding Of Two-Channel Audio Signals By Means Of Complex-Valued Stereo Prediction,” in Proc. IEEE ICASSP-2011, pp. 497-500, 2011.

[7] Christof Faller, and Frank Baumgarte, “Binaural Cue Coding: A Novel and Efficient Representation of Spatial Audio,” in Proc. IEEE ICASSP-2002, vol: 2, pp. II-1841 - II-1844, 2002.

[8] F. Baumgarte, and C. Faller,“Binaural Cue Coding-part I : Psychoacoustic fundamentals and Design Principles,” in IEEE Trans. on Speech and Audio Proc., vol. 11, No. 6, pp. 509-519, June 2003.

[9] F. Baumgarte, and C. Faller,“Binaural cue coding-part II : Schemes and applications,” in IEEE Trans. on Speech and Audio Proc., vol. 11, No. 6, pp. 520-531, June 2003.

[10] C. Faller, “Parametric Multichannel Audio Coding: Synthesis of Coherence Cues," IEEE Trans. Speech and Audio Proc., vol. 14, No. 1, pp. 1-12, Jan. 2006.

[11] J. Breebaart, et al.,“Parametric Coding of Stereo Audio,” in EURASIP Journal on Applied Signal Processing, vol 2005, No. 9, pp 1305 - 1322, June 2005.

[12] A. Kohlrausch, “Auditory filter shape derived from binaural masking experiments," J. Acous. Soc. America, vol. 84, no. 2, pp. 573-583, 1988. 16

[13] B. R. Glasberg and B.C.J. Moore, “Derivation of auditory filter shapes from notched-noise data," Hearing Research, vol. 47, no. 1-2, pp . 103-138, 1990.

[14] K. Suresh, and T. V. Sreenivas, “MDCT Domain Analysis and Synthesis of Reverberation for Parametric Stereo Audio,” in AES 123th Convention, 2007 October 5-8, New York.

[15] K. Suresh, and T. V. Sreenivas, “Parametric stereo coder with only MDCT domain computations,” IEEE International Symposium on Signal Processing and Information Technology, pp. 61-64, December 2009.

[16] K. Suresh and T. V. Sreenivas, “Linear Filtering in DCT-IV/DST-IV and MDCT/MDST Domain”, Signal Processing, vol 89, Issue 6, pp 1081-1089, June 2009.

[17] T. Painter, and A. Spanias, “Perceptual Coding of Digital Audio", Proc. IEEE, vol. 88, no 4, pp. 451-513, 2000.

[18] K Suresh and T. V. Sreenivas, “Direct MDCT Domain Psychoacoustic Modeling”, IEEE International Symposium on Signal Processing and Information Technology, pp. 742-747, December 2007.

[19] ITU/ITU-R BS 1534. Method for subjective assessment of intermediate quality level of coding systems, 2001.