# PARAMETRIC CODING OF STEREO AUDIO BASED ON PRINCIPAL COMPONENT ANALYSIS

*Manuel Briand, David Virette*

Speech and Sound Technologies and Processing (SSTP) Laboratory
France Telecom R&D, Lannion, France
{manuel.briand | david.virette}
@orange-ft.com

*Nadine Martin*

Image and Signal processing Laboratory (LIS)
CNRS UMR 5083
INP-Grenoble, France
nadine.martin@lis.inpg.fr

## ABSTRACT

Low bit rate parametric coding of multichannel audio is mainly based on Binaural Cue Coding (BCC). Another multichannel audio processing method called upmix can also be used to deliver multichannel audio, typically 5.1 signals, at low data rates. More precisely, we focus on existing upmix method based on Principal Component Analysis (PCA). This PCA-based upmix method aims at blindly create a realistic multichannel output signal while BCC scheme aims at perceptually restitute the original multichannel audio signal. PCA-based upmix method and BCC scheme both use spatial parameters extracted from stereo channels to generate auditory events with correct spatial attributes *i.e.* sound sources positions and spatial impression. In this paper, we expose a multichannel audio model based on PCA which allows a parametric representation of multichannel audio. Considering stereo audio, signals resulting from PCA can be represented as a principal component, corresponding to directional sources, and one remaining signal, corresponding to ambience signals, which are both related to original input with PCA transformation parameters. We apply the analysis results to propose a new parametric coding method of stereo audio based on subband PCA processing. The quantization of spatial and energetic parameters is presented and then associated with a state-of-the-art monophonic coder in order to derive subjective listening test results.

## 1. INTRODUCTION

Multichannel audio has been established in the consumer environment through the success of DVD-Video players for home theater systems. Moreover, the streaming technology over IP used as a broadcast service is requesting multichannel audio at low data rates. Therefore, multichannel audio coding and processing methods have been investigated by many researchers during last decade.

The first method is denoted as multichannel audio coding. Matrix surround coding schemes and parametric audio coding schemes are the two main multichannel audio coding techniques currently used. Matrix surround coding scheme such as Dolby Pro Logic [1] consists in matrixing the channels of the original multichannel signal in order to reduce the number of signals to be transmitted. Nevertheless, this multichannel audio coding method cannot deliver high quality (close to transparency) at low data rates. That is made possible with low bit rate parametric audio coding mainly based on Binaural Cue Coding (BCC) [2]. This coding scheme represents multichannel audio signals by one or several downmixed audio channels plus spatial cues extracted from the original channels. The spatial cues refer to the auditory localization cues defined as interaural time and level differences (ITD and ILD) extracted from input channel pairs in a subband domain and then denoted as inter-channel time and level differences (ICLD and ICTD). BCC uses filterbanks with subbands of bandwidths equal to two times the equivalent rectangular bandwidth (ERB) defined in [3]. Moreover, the inter-channel coherence (ICC) is also extracted in order to recre-ate the diffuseness of the original multichannel input. Indeed, the multichannel audio synthesis at the decoder side is based on the ICC parameter which yields a coherence synthesis relying on late reverberation. The downmixed audio channel (in case of mono downmix) is decoded and then filtered by late reverberation filters which deliver several decorrelated audio channels (see [2] for more details). Then, these signals are combined according to spatial cues (ICTD, ICLD and ICC) such that the ICC cues between the output subbands approximate those of the original audio signal. Then, BCC scheme achieves a drastically data rate reduction by transmitting a perceptually encoded downmixed signal plus quantized spatial cues. Moreover, BCC scheme achieves a better audio quality and perceived spatial image than matrix surround coding scheme (see subjective tests results in [2]). From a spatial attribute point of view, BCC synthesis restricted to ICLD and ICTD achieves the desired sources positions and coloration effects caused by early reflections but suffers from auditory spatial image width reduction. Indeed, spatial impression is related to the nature of reflections that occur following the direct sound. Then, BCC synthesis based on late reverberation mimicks different reverberation times and then achieves a better spatial impression closer to the original multichannel input.

A second multichannel audio processing method, called upmix, classically converts the existing stereo audio contents into five-channel audio compatible with home theater systems. So, the decoding process of BCC has the common intention with upmix method which is to deliver multichannel audio signal – considering upmix stereo input and BCC stereo downmix. *A priori*, more information is available for BCC scheme *i.e.* spatial parameters, than for blind upmix method. However, upmix method uses the spatial characteristics and the coherence of the stereo signal to synthesize a mul-tichannel audio signal, typically 5.1 signal with rear channels considered as ambience channels – defined as diffuse surround sounds – and a center front channel corresponding to the sources panned across the original stereo channels. More precisely, we focus on existing PCA-based upmix method described in [4, 5]. The first step of the upmix algorithm in [4] consists in a Principal Component Analysis (PCA) of the stereo signal. A

subband processing of this upmix method has been recently proposed in [5]. The PCA is equivalent to a rotation of the stereo signal coordinate system and results in one principal component signal and one remaining signal. The principal component signal corresponds to the dominant sources present in the original stereo. Then, the center channel results from the weighting of this principal component by a coefficient derived from the rotation angle of the coordinate system. The rear channels result from the weighting of the remaining signal by a coefficient derived from the correlation coefficient of the stereo signal.

Starting from PCA approach, we propose a general model that may be applied both to parametric representation of multichannel audio signals and upmix methods. Moreover, we apply the analysis results to propose a new parametric audio coding method based on frequency subband PCA processing. This paper is organized as follows. A general model of multichannel audio signals is presented in section 2. Then, subband PCA of stereo audio is considered and results in an energy compaction and a parametric representation related to original stereo with PCA transformation parameters. Finally, a new parametric coding method of stereo signals is presented in section 3. The encoding stage focuses on spatial and energy parameter extraction and quantization while the decoding stage addresses the parametric synthesis of ambience signal in order to achieve accurate inverse PCA. Subjective listening test results are presented in section 4 in order to evaluate the performance of the parametric stereo coding method.

## 2. PARAMETRIC REPRESENTATION OF MULTICHANNEL AUDIO BASED ON PCA

Multichannel audio signals either originate from studio (artificial) produced signals or from live (natural) recorded signals [6]. Live recording involves many different setups and microphones types which determine the amount of interferences and reverberation on each channel. Decorrelated reverberant channels yield a sensation of realistic ambience. Ambience has a complex audio content, perceptually perceived in background and very heterogeneous. Ambience could be defined as "the sound of the place in which sources are". Such an audio content includes acoustic effects of reverberant volumes and reflective features plus the background *i.e.* the acoustic accumulation of many small sources that are not the identified sources of interest; for example, audience noise. In studio recording, sound sources (instruments) are individually recorded and then processed. The processing of recorded sound sources consists in applying panning functions to the sound sources and then mixing them with synthetic reverberation. In order to increase the perception of spaciousness, weakly correlated reverberation impulse responses are used.

Under these assumptions, we define a general model for multichannel audio signals constituted of $M$ channels. We assume the presence of $D$ directional sources that can be easily localized. These directional sources are distributed over the $M$ channels by means of panning laws. Moreover, we consider $M$ ambience signals *i.e.* one by channel, which are defined as secondary sources and decorrelated reverberant components of directional sources *i.e.* room effect. Naturally, these ambience signals are decorrelated from one channel to another and weakly correlated with directional sources. The model of each channel is defined as the sum of directional sources, weighted according to their spatial perceived positions, and one ambience signal. Signal channels following this instantaneous mixture model are strongly correlated with the presence of directional sources among several channels. So, the time domain multichannel signal $\mathbf{C_M} = (C_1, \cdots, C_M)^T$ ($^T$ denotes matrix transposition) can be written as:

$$C_m(t) = \sum_{d=1}^{D} [g_{m,d}(t).S_d(t)] + A(t) \qquad (1)$$

where $m \in [1, \cdots, M]$ and $g_{m,d}(t)$ are the panning functions (gains) applied to the directional sources $\mathbf{S_D} = (S_1, \cdots, S_D)^T$ of the $m^{\text{th}}$ channel. Finally, room effects on directional sources are distributed over all channels because these decorrelated components belong to the ambience signals $\mathbf{A_M} = (A_1, \cdots, A_M)^T$.

In a general audio coding context, the separation of dominant sources (even mixed with part of the overall ambience) from background ambiences could be seen as a pre-processing step achieved with PCA before applying a dedicated coding scheme. Indeed, the transmission of encoded dominant sources already provides a basic audio scene of the original input. If the coding method foresees the generation of an ambience signal, then, the decoding process can achieve better spatial impression.

PCA also known as Karhunen-Loève Transform (KLT) has been used by numerous researchers for miscellaneous applications. D.T. Yang, C. Kyriakakis and C.-C. Jay Kuo use KLT as an interchannel redundancy removal in a multichannel audio coding context in [7]. Indeed, PCA/KLT achieves an energy concentration over PCA outputs according to the distribution of eigenvalues. Moreover, PCA is theoretically considered as the optimal decorrelation method *i.e.* the covariance matrix of PCA output signals is diagonal. KLT is only optimal in the case of stationary signals, which is an admitted assumption for signal processing with small block length (typically 20 ms for audio/speech codecs). The authors in [7] propose a modified Advanced Audio Coding encoder which performs better coding gain using KLT applied to the multichannel input.

From the audio model defined by equation (1), our intention is to characterize the PCA outputs and then achieve a parametric coding of this compact representation of multichannel audio.

Eigenvalues distribution of such multichannel audio signals precisely indicates how the original audio content is distributed over the PCA outputs. An analysis of estimated eigenvalues from a synthetic stereo signal made up of real recordings has been achieved in [8]. This analysis has shown that the highest eigenvalue power corresponds to the power of the dominant directional source. The lowest eigenvalue power corresponds to the power of the secondary directional sources plus the power of the overall ambience. Moreover, a comparison of time-domain estimated eigenvalues and frequency subband estimated eigenvalues has shown that subband analysis results in a better discrimination of directional sources. Indeed, some directional sources considered as secondary sources with the time domain analysis can be considered as dominant directional sources with the subband analysis when these sources have different frequency support. Finally, with a subband analysis, the lowest eigenvalue has a power much closer to the original ambience signal mean power and respectively, the highest eigenvalue has a power closer to the sum of directional source powers.

This energy concentration over eigenvalues can also be computed directly from the output signals resulting from PCA. These signals result from a rotation of the stereo channels (see [8] for more details) and are then denoted as principal component signal ($PC$) related to the highest eigenvalue and ambience signal

($A$) related to the lowest eigenvalue. Time-domain PCA processing and frequency subband (following ERB scale) PCA processing are presented in [8]. Then, a relevant measurement of the energy compaction into $PC$ is achieved by estimating the Principal Component to Ambience energy Ratio $PCAR[n] = 10\log_{10}\left(\sum PC[n]^2 / \sum A[n]^2\right)$ in dB, estimated from the $N$ signal samples resulting from time-domain or subband rotation(s) of the stereo analyzed block $n$. For homogeneity, the $PCAR$ is estimated from $PC$ and $A$ signal blocks which have the same length ($N$) than the processed stereo channel blocks. Then, the $PCAR$ is function of the length of the stereo processed blocks. The Figure 1 is addressing a comparison of the average $PCAR$ over analyzed blocks of length $N = 1024$, 2048 and 4096 samples. These energy ratios have been estimated over a stereo signal corpus.

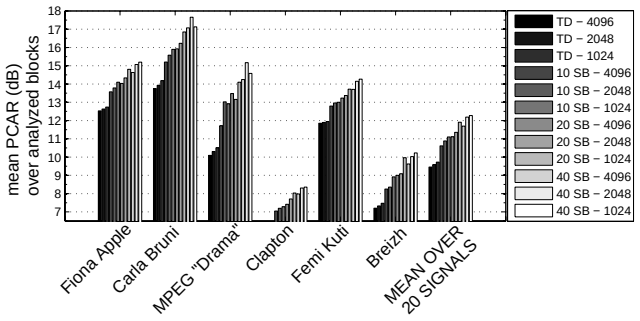


Figure 1: *From time-domain (TD) and frequency subband (SB) PCA outputs, PCAR are estimated with the following block processing lengths: 1024, 2048 and 4096 samples. Mean $PCAR$ over all blocks from 6 stereo signals and mean $PCAR$ over the stereo signal corpus are presented.*

Moreover, $PCAR$ estimated from subband rotated signals is function of both time and spectral resolutions. In fact, the frequency resolution is determined by the frequency scale used for subband separation and the number of frequency subbands. $PCAR$ estimated from time-domain PCA output signals are also compared to $PCAR$ estimated from subband PCA output signals (see Figure 1) with $N_b = 10$, 20 and 40 subbands of bandwidth following the ERB scale. The energy compaction into the principal component is higher with a subband analysis comparing to time-domain analysis (see Figure 1, mean difference about 2 dB) and naturally increasing with the number of subbands (maximum difference about 5 dB for the MPEG "Drama" sample).

## 3. PARAMETRIC CODING OF STEREO AUDIO

Starting from the analysis made in section 2, it is possible to encode a stereo signal with frequency subband PCA pre-processing. PCA is used as a power concentration processing such as the Mid-/Side coding scheme which encodes the sum and difference signals of stereo channels. The principal component and transformation parameters should be encoded and transmitted in order to recover the main information of the original input. To obtain good performances in the inverse transformation process, the ambience component should be transmitted with a bit rate compatible with the desired audio quality level. Even if PCA features, such as power concentration and full decorrelation are required for optimal bite rate reduction, traditional encoding of the transformed signals does not yield significant coding gain (see [9]). In order to provide a low bit rate audio coding method compatible with bit rate constrained applications, a parametric coding of the ambience signal is achieved.

### 3.1. Parameters based on PCA transformation and ambience energy

We consider band limited signals in the frequency domain. The frequency transform applied to stereo channels is the short time Fourier transform (STFT). The parameters of the STFT used are a sine window of length equal to $N = 4096$ samples, the transform size is also equal to $K = 4096$ (no zero-padding) and the frame overlap is 50%. Then, a $N_b = 20$ subband rectangular frequency windowing, following the ERB scale, is applied to the channel ($C$) complex spectra $F_C[n,k] = |F_C[n,k]|e^{j\Phi_C[n,k]}$, with frequency index $k$ such as $k \leq f_s/2$, where $f_s$ is the sampling frequency. This process results in $N_b$ frequency subband spectra. The frequency bins of subband $b$ for each frame $n$ belong to the interval $(k_b, \cdots, k_{b+1}-1)$, then $F_C^b[n,k] = F_C[n,k_b], \cdots, F_C[n,k_{b+1}-1]$.

#### *3.1.1. Subband parameter extraction*

The first parameter which needs to be transmitted to the decoder is related to the PCA transformation. As described in [8], PCA of stereo channels is obtained by rotating the stereo subband channels over the stereo covariance eigenvector basis. This subband rotation is performed for each frame $n$ and subband $b$ according to the following expression of the estimated rotation angle $\theta[n,b] \in [0;\pi/2]$

$$\begin{aligned}\theta[n,b] &= \frac{1}{2}\tan^{-1}\left[\frac{2\times|R_{12}[n,b]|}{R_{11}[n,b]-R_{22}[n,b]}\right] \\ &\quad + \begin{cases} 0, \text{ if } R_{11}[n,b]-R_{22}[n,b] \geq 0 \\ \frac{\pi}{2}, \text{ else} \end{cases}\end{aligned} \quad (2)$$

where the stereo subband covariance matrix $\mathbf{R}$ is estimated from the subband spectra $F_{\overline{L}}[n,k]$ and $F_{\overline{R}}[n,k]$ of the windowed and centred signals $\overline{L}[n] = L[n] - E[L[n]]$ and respectively for $\overline{R}[n]$. The channel auto-covariance ($R_{11}$ and $R_{22}$) correspond to the mean power spectral density of the subband spectra and the channel cross-covariance ($R_{12}$) is estimated from the cross-spectrum of the stereo channels

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}, \text{with}$$

$$\begin{cases} R_{11}[n,b] = \frac{2}{NK}\sum_{k=k_b}^{k_{b+1}-1}|F_{\overline{L}}[n,k]|^2 \\ R_{12}[n,b] = \frac{2}{NK}\Re\left(\sum_{k=k_b}^{k_{b+1}-1}F_{\overline{L}}[n,k]\cdot F_{\overline{R}^*}[n,k]\right) \\ R_{22}[n,b] = \frac{2}{NK}\sum_{k=k_b}^{k_{b+1}-1}|F_{\overline{R}}[n,k]|^2 \end{cases} \quad (3)$$

A robust and effective estimation of the rotation angle needed to achieve PCA/KLT is expressed by the equation (3) which permits an estimation only based on the covariance matrix elements. Moreover, the physical meaning of this rotation angle can be interpreted as the perceived position of the dominant directional source

localized in the stereo image $[-30, 30]^o$. In fact, this rotation angle is considered as a panning angle in the PCA-based upmix method to compute panning matrix which can generate realistic multichannel signals from PCA outputs. These subband rotation angles are then used to rotate the subband input stereo data blocks. Subband PCA processing results in two frequency components for each subband: the principal component $F^b_{PC}[n,k]$ and the ambience component $F^b_A[n,k]$.

Then, these subband signals resulting from PCA are analyzed and energetic parameters are extracted in order to be able to generate the ambience signal at the decoder side. Actually, the ambience synthesis should be achieved only from the decoded principal component and these energetic parameters. The parametric coding method does not consider the phase of the ambience signal. According to the subjective comparison test achieved in [8], the only perceptually relevant parameter of the ambience signal which needs to be transmitted is related to the ambience signal energy. Therefore, $PCAR$ is estimated from subband spectra from each analyzed block and each subband according to following expression

$$PCAR[n,b] = 10\log_{10}\left(\frac{\sum_{k=k_b}^{k_{b+1}-1}|F_{PC}[n,k|^2}{\sum_{k=k_b}^{k_{b+1}-1}|F_A[n,k|^2}\right) \quad (4)$$

Finally, the PCA-based coding scheme applied to stereo signal is presented on Figure 2. The time domain signal PC is generated with the (overlap-add) sum of all synthesized block signals obtained with inverse short time Fourier transform (STFT$^{-1}$) of the sum of all frequency subband components. After subband PCA processing, the extracted parameters are quantized and perceptual monophonic coding of the principal component signal is achieved.

### *3.1.2. Subband parameter quantization*

Our intention is to realize a low complexity uniform quantization of the subband parameters and then estimate the mean bit rate with Huffman coding. A training basis has been processed in order to estimate the dynamics and the precision needed for the quantization of subband parameters. Subband rotation angles and PCAR have been extracted from a training basis made up of miscellaneous stereo recordings with a total duration of about one hour. The stereo channel blocks have been processed by the frequency transform described at the beginning of section 3.1. Parameters have been estimated according to equations (2) and (4).

Concerning the estimated rotation angle varying in the interval $[0; 90]^o$, we propose the quantization of the mean value of subband rotation angles for each analyzed block $n$, $\overline{\theta}$, defined as

$$\overline{\theta}[n] = \frac{1}{N_b}\sum_{b=1}^{N_b}\theta[n,b] \quad (5)$$

Mean rotation angle transmission assures a basic spatial reconstruction at low data rate with a reduced spatial image width. Then, with an additional data rate corresponding to the subband rotation angles, the original spatial image width can be recovered. Actually, we propose the quantization of differential subband rotation angles $\theta_D[n,b] = \theta[n,b] - \theta[n,b-1]$, if $b > 1$. Differential subband rotation angles $\theta_D$ have pickier distribution with dynamics (in degrees) decreasing along frequencies. Then, a better coding gain can be achieved. The first subband rotation angle is processed independently and its mean removed value $\theta_{MR}[n,1] = \theta[n,1] - \overline{\theta}[n]$ is also considered for the quantization process.

Considering the estimated $PCAR$, we assume that the ambience signal has globally low energy level which decreases along frequencies. As a result, we also propose the quantization of the differential subband $PCAR$, $PCAR_D[n,b] = PCAR[n,b] - PCAR[n,b-1]$ if $b > 1$.

These parameters are then uniformly quantized. According to perceptual criteria, the quantization process aims at introducing quantization errors which are just inaudible.

For the rotation angles, the minimum audible angle (MAA), as described in [10], is considered. The performance in localization is a complex function which depends on the nature of the stimulus and also on the stimulus frequency. For instance, sinusoidal signals can be discriminated with $\pm 1^o$ precision for a source at $0^o$ position (direct front of the listener) and this only for frequencies lower than 1 kHz. More precisely, the localization accuracy is decreasing when the sound source is moving away from the direct front of the listener and also when the frequency of the sound source is increasing. For example, a sinusoidal signal at $30^o$ position (on the left side of a stereo setup) can be discriminated with at least $\pm 5^o$ precision for stimulus frequencies higher than 2 kHz, still extracted from [10]. So, considering real sound sources with spatial position belonging to the stereo setup interval $[-30; 30]^o$, the corresponding precision for the mean rotation angle $\overline{\theta}$ varying in the interval $[0; 90]^o$ is chosen equal to at worst $\pm 3^o$. The first subband mean removed rotation angle $\theta_{MR}$ and the differential subband rotation angles $\theta_D$ values are varying in different intervals with a dynamics decreasing along frequencies. Moreover, these modified subband rotation angles do not need the same precision as the vital mean rotation angle; they only aim at refining the spatial position of the sound sources. We propose to attribute different level of precision for these mean removed and differential subband angles. Therefore, considering the subband differential rotation angles $\theta_D$, we follow the academic results presented in [10] with a precision decreasing when frequency increases:

- $\pm 4^o$ with frequency: $f < 1$ kHz,
- $\pm 7.5^o$ with frequency: $f < 5$ kHz,
- $\pm 10^o$ with frequency: 5 kHz $< f < fs$.

Considering the $PCAR$ expressed in decibels (dB) as the energy difference between the principal component and the ambience signal, audible difference between reconstructed and original stereo signals are avoided with an energetic precision about $\pm 3$ dB. More precisely, the precision allocated to the original $PCAR$ of the first subband is slightly higher than the precision allocated to the subband differential $PCAR_D$. The $PCAR$ quantization step is less critical than just noticeable differences usually used by frequency quantizers because $PCAR$ is used to synthesize the ambiance spectral envelope.

The precision allocated for each parameter determines the amount of bits needed by the quantization processes. This quantization step is followed by a Huffman coding which achieves a reduction of the mean code length and then, delivers a mean bit rate of this parametric stereo coding method. From a training basis, this parameter mean bit rate has been estimated and compared with parameter mean bit rates estimated from other well known stereo signals extracted from MPEG audio basis. The resulting mean bit rates estimated from the MPEG audio basis are on average (2.8 kbps) slightly lower than the mean bit rate estimated from

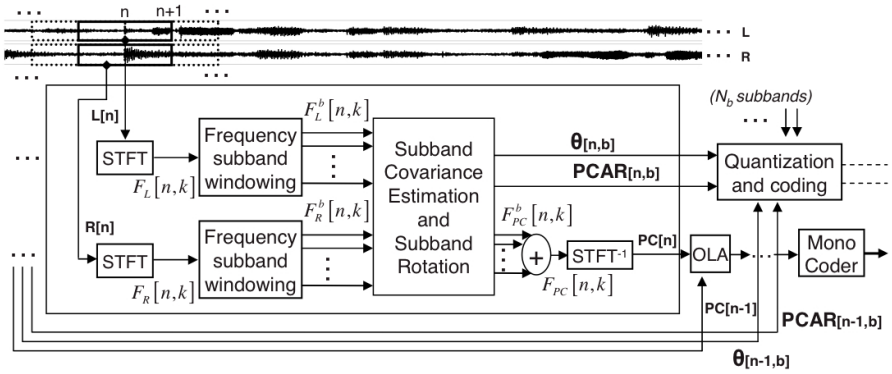


Figure 2: *Parametric coding of stereo signals based on subband frequency PCA processing.*

the training basis (3 kbps). So, we can conclude that the quantization process is weakly sensitive to the training basis.

In the final implementation of the parametric stereo coder, the quantized values of the subband rotation angles are used to achieve subband PCA. Afterwards, subband $PCAR$ are extracted from the PCA outputs, quantized and transmitted with the quantized rotation angles. In order to reduce even more the bit rate, a joint polar quantization of the rotation angles and $PCAR$ could be considered.

### 3.2. Ambience synthesis to achieve inverse PCA

The decoding scheme is based on the generation of an ambience signal $A'$, from the decoded signal $PC'$ and the dequantized parameters (see Figure 3). Therefore, the inverse PCA can then synthesize a stereo signal perceptually as close as possible from the original stereo. Due to the PCA property of decorrelation, the decoder should generate an ambience signal weakly correlated to the decoded principal component. However, the frequency synthesis of subband signal $A'$ from the decoded principal component $PC'$ and the energy parameters $PCAR^Q[n,b]$, only provide $A'$ spectral envelope. To achieve weak correlation between $PC'$ and $A'$, we propose the use of random phase all-pass filters as described in [11]. We perform a subband filtering (filter $H$ on Figure 3) of $F^b_{A'}[n,k]$ spectra.

So, the decoder can operate the inverse subband PCA from the signal subband spectra $F^b_{PC'}[n,k]$ and $F^b_{A'_H}[n,k]$ and the dequantized rotation angles $\theta^Q[n,b]$ (see Figure 3). Afterwards, the sum of all subband signals, obtained from inverse subband PCA, is transformed to time domain by inverse STFT to generate the stereo signal $(L', R')$.

## 4. SUBJECTIVE LISTENING TEST

In order to evaluate the PCA based parametric stereo coding method, a subjective listening test has been conducted. This subjective test aims at comparing the PCA-based parametric stereo coding method to the parametric stereo coding method used in the state-of-the-art High-Efficiency Advanced Audio Coding Parametric Stereo (HE-AACv2). Then, to achieve accurate comparison, several stereo signals encoded by HE-AACv2 at 24 kbps are perceptually compared to stereo signals processed by the PCA-based parametric stereo coding method described in section 3. More precisely, the principal component signal is encoded by the monophonic profile of the HE-AAC encoder at 22 kbps and the quantized parameters are transmitted at a bit rate around 3 kbps. A lower bit rate of the HE-AAC mono profile would have generated monophonic signal with lower bandwidth than the bandwidth (16 kHz) of the stereo signals generated by the HE-AACv2 operating at 24 kbps. Then, the overall bit rate for the PCA-based parametric stereo coding method (PCA-HE-AAC) is about 25 kbps using 20 frequency subbands and a parameter update rate linked to the 4096 sample block length.

A subjective listening test following the MUSHRA methodology has been conducted. Thirteen listeners participated in this headphone listening experiment. All listeners were instructed to evaluate both the spatial audio quality as well as other audio coding artefacts. Listeners had to rate the perceived quality of seven processed items against the original stereo on a 100-point scale with two anchors corresponding to 3.5 kHz and 7 kHz low pass filtered version of the original stereo. Five original audio items were extracted from the MPEG audio basis and the other two stereo signals were chosen for their impressive stereo image. A comparison of the MUSHRA scores is presented on Figure 4-(a)-(b).

The stereo items processed by the HE-AACv2 have slightly higher score than the stereo audio processed by the PCA-HE-AAC (see Figure 4-(a)). This small difference can be explained by the fact that the PCA-HE-AAC codec uses a fixed and higher parameter update rate (4096 samples) than the HE-ACCv2, which results in degraded transients; this is the case for the stereo item “Guitar+Castanets".

Although, the time-frequency resolution needs to be considered and improved to avoid coding artefacts, the spatial impression given by stereo signals processed by the PCA-based parametric coder is very close to the original stereo image. The spatial synthesis of the PCA-based parametric coder is the main advantage of this coding method and has been considered, by the listeners, as more stable than the spatial impression delivered by the HE-AACv2 coder. This explains that on average, there is no significant difference between both parametric stereo coding methods (see Figure 4-(b)).

## 5. CONCLUSIONS

In this paper, an instantaneous mixture model for multichannel audio is defined as the sum of weighted directional sources and ambiences. Then, PCA of stereo signals following the model has been considered. $PCAR$ estimation has shown how a frequency

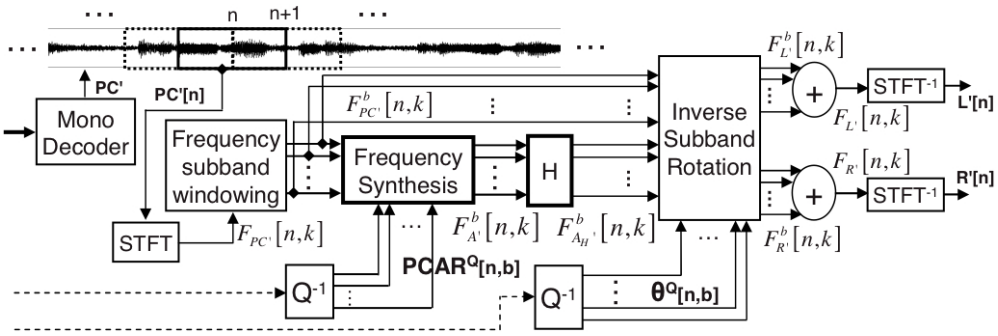


Figure 3: *Parametric decoding of stereo signals based on inverse subband PCA processing.*

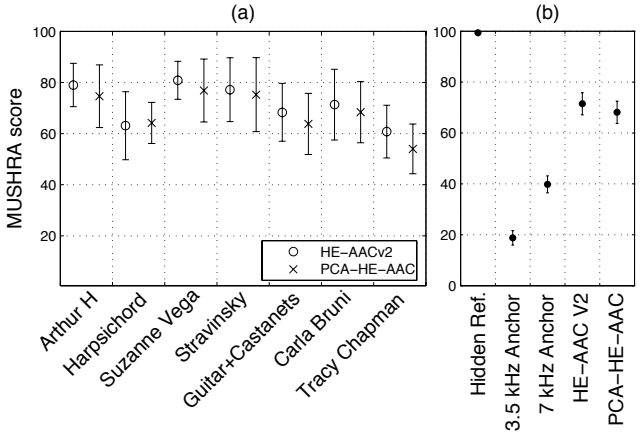


Figure 4: *(a)-MUSHRA mean gradings and* 95% *confidence intervals over all listeners as a function of test item and the parametric stereo coding method. (b)-MUSHRA mean gradings and* 95% *confidence intervals over all listeners and all items.*

subband PCA yields more efficient power concentration than classical time-domain PCA. Stereo audio is represented as a principal component, related to directional sources, plus a less energetic ambience signal, related to decorrelated ambiences, which provides spatial impression.

The main point addresses a parametric coding method of stereo audio based on subband PCA output signals. A stereo signal can be represented as a principal component and rotation angles which both permit a basic audio stereo reconstruction. Moreover, the ambience signal, related to the spatial impression components, has globally low energy level and no perceptually relevant phase information. Therefore, this ambience signal allows a parametric coding of stereo audio by means of PCA rotation angles and ambience energy parameters. To achieve better coding gain, subband differential parameters are uniformly quantized according to perceptual criteria. Then, Huffman coding of the quantized parameters is achieved in order to estimate a mean bit rate (about 3 kbps) of the parametric stereo coding method. The analysis has showed that the parameter quantization is not sensitive to a training basis.

Finally, the conducted listening test yield to the conclusion that there is no perceptually significant difference comparing the HE-AACv2 codec and the parametric stereo coding method based on PCA associated with the monophonic profile of the HE-AAC codec. The parameter update rate should be adapted to the audio content to improve the audio quality.

## 6. REFERENCES


[1] R. Dressler, "Dolby surround Prologic II decoder: Principles of operation," Dolby Laboratories, Tech. Rep., 2000, [Online] http://www.dolby.com/assets/pdf/tech_library/209_Dolby_Surround_Pro_Logic_II_Decoder_Principles_of_Operation.pdf.

[2] C. Faller, "Parametric coding of spatial audio," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, July 2004, thesis No. 3062, [Online] http://library.epfl.ch/theses/?nr=3062.

[3] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103–138, 1990.

[4] R. Irwan and R. M. Aarts, "Two-to-five channel sound processing," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 914–926, Nov. 2002.

[5] Y. Li and P. F. Driessen, "An unsupervised adaptative filtering approach of 2-to-5 channel upmix," in *119th Conv. Audio Eng. Soc.,* New York, USA, Oct. 2005, preprint 6611.

[6] Producers & Engineers Wing Surround Sound Recommendations Committee, "Recommendations for surround sound production," The National Academy of Recording Arts & Sciences, Tech. Rep., 2004.

[7] D. T. Yang, C. Kyriakakis, and C. C. Jay Kuo, *High-Fidelity Multichannel Audio Coding*. EURASIP Book Series on Sig. Proc. and Communications, 2004.

[8] M. Briand, D. Virette, and N. Martin, "Parametric representation of multichannel audio based on Principal Component Analysis," in *120th Conv. Audio Eng. Soc.,* Paris, France, May 2006, preprint 6813.

[9] R. G. van der Waal and R. N. J. Veldhuis, "Subband coding of stereophonic digital audio signals," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP'91),* Toronto, Canada, May 1991, pp. 3601–3604.

[10] B. C. J. Moore, *Psychology of Hearing*. Academic Press, 1993.

[11] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *117th Conv. Audio Eng. Soc.,* San Francisco, USA, Oct. 2004, preprint 6291.