A New Approach for Low-Delay Joint-Stereo Coding

Hauke Krüger and Peter Vary

Institut für Nachrichtengeräte und Datenverarbeitung (IND), RWTH Aachen, 52056 Aachen
E-Mail: {krueger,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

In this contribution, a new approach for the coding of stereophonic audio signals based on inter-channel linear prediction is proposed. In contrast to other recent contributions on joint-stereo coding where left-to-right- and/or right-to-left-channel linear prediction is used, in our approach each of the two channels is predicted by filtering the center stereo image (sum) of both channels. The optimal prediction coefficients for both channels can be computed in a way very similar to single-channel linear prediction, and it turns out that the proposed technique is a generalization of Mid/Side (M/S) and Left/Right (L/R) joint-stereo coding. Since the new approach is based on a time domain representation of the signals, it is especially well suited for stereo coding with low algorithmic delay. Due to its modularity, it is also suitable to extend any existing monaural speech or audio codec toward stereo functionality.

1 Introduction

In the history of stereo audio transmission, in Frequency Modulated (FM) radio, broadcasting of stereophonic signals started already in 1961. The basis for FM stereo broadcasting is the production of a mid and a side channel signal (M/S stereo) from the left and right channel signals. In each modulated FM radio channel, the mid channel signal is transmitted in the baseband spectrum and the side channel signal in the spectrum related to the amplitude modulated *double-sideband suppressed carrier signal* (DSSCS) [6][3]. Still nowadays, FM radio receivers may reconstruct either only the monaural mid channel representation (mono) of the input stereo signal from only the baseband spectrum, or the complete stereo image signal if also the DSSCS signal is demodulated. In digital audio compression, a lot of confusion is related to the term joint-stereo coding. In the literature, it is referred to as both, M/S and Intensity Stereo coding. The target of joint-stereo coding is to enable a higher compression ratio in a joint coding approach in comparison to an approach in which the signals for left and right channel are coded independently.

A lot of joint-stereo approaches in the literature are based on a high resolution frequency domain representation of the input signal (e.g. Intensity Stereo Coding, [2],[5]) and therefore related to a high algorithmic delay. In contrast to these techniques, joint-stereo coding approaches in the time domain better achieve low algorithmic delay. In [4], an adaptive inter-channel predictor is proposed that is composed of an inter-channel FIR prediction filter and a delay. Predictor filter coefficients and inter-channel delay adapt to the given signals for left and right channel. The target of this approach is to produce an estimate of the first channel on the basis of the second channel to reduce the signal variance of the predicted channel and hence save bits. Adaptive multichannel prediction is also investigated in [8] and revisited in [1]. In this case, inter- and intra-channel predictors are optimized in a joint way to produce residual signals with reduced signal variance in both channels to reduce the overall bit rate for lossless coding. Both techniques are not suitable to extend existing mono codecs in a hierarchical way.

In this paper, an alternative approach for joint-stereo coding is proposed. It operates in the time domain and enables low algorithmic delay. Compared to the approaches given in the literature, in our new approach, the sum of left and right stereo input signal (the mono representation) is filtered by linear phase FIR filters to predict the left *and* the right channel. Due to its modularity, the new approach is suitable to extend existing monaural codecs toward the coding of stereo signals while preserving backwards compatibility with monaural transmission.

2 M/S and L/R Joint-Stereo Coding

The principle of Mid/Side (M/S) joint-stereo coding is shown in Figure 1. Given the signals of the left and the right audio channel as $x_R(k)$ and $x_L(k)$ respectively, the mid and the side channel signals $x_M(k)$ and $x_S(k)$ are calculated in the encoder as

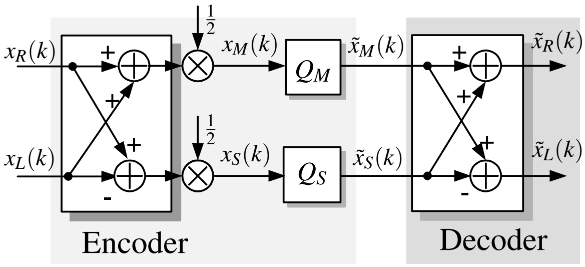


Figure 1: Principle of M/S joint-stereo coding.

$$x_M(k) = (x_R(k) + x_L(k))/2 \quad (1)$$
$$x_S(k) = (x_R(k) - x_L(k))/2. \quad (2)$$

Both signals are quantized in independent quantizers, Q_M and Q_S respectively, and transmitted to the decoder. The quantized left and right channel signals are reconstructed from the quantized versions of the mid and the side channel signal as

$$\tilde{x}_R(k) = \tilde{x}_M(k) + \tilde{x}_S(k) \quad (3)$$
$$\tilde{x}_L(k) = \tilde{x}_M(k) - \tilde{x}_S(k). \quad (4)$$

In a typical audio signal recording, often, a strong mid channel signal component is present so that the signal variance of $x_M(k)$ is significantly higher than that of $x_S(k)$ which can be exploited to reduce the overall bit rate compared to independent quantization of left and right channel. M/S joint-stereo coding is used in a fullband approach in Figure 1 but can also be applied to subband signals produced by a filterbank [7].

In the presence of signals with a very dominant signal component in one channel, M/S coding does not provide any coding advantage. In this case, L/R joint-stereo coding achieves a bit rate reduction if more bit rate is allocated for the channel with the dominant signal component than for the other channel. Switching between M/S and L/R coding, however, must be signaled to the decoder.

3 The New Approach

Our new approach operates in the time domain to achieve low algorithmic delay and is shown in Figure 2. From the right and the left channel input signal, in the first step a mono signal is calculated,

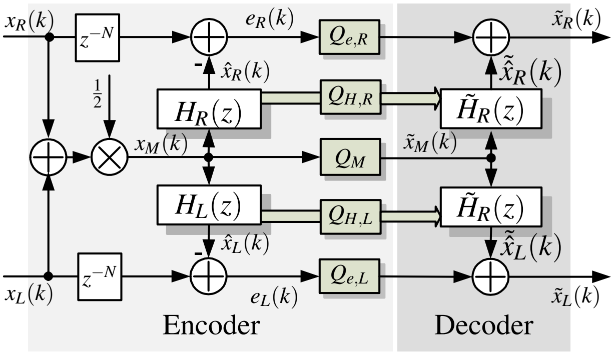


Figure 2: New approach for joint-stereo coding.

$$x_M(k) = \frac{x_R(k) + x_L(k)}{2}. \qquad (5)$$

The signals $\hat{x}_L(k)$ and $\hat{x}_R(k)$ are produced as the estimate for the left and right channel input signals by means of linear filtering of the mono signal with system functions $H_L(z)$ and $H_R(z)$ respectively. The filters are symmetric linear phase FIR filters with $(2 \cdot N + 1)$ filter coefficients,

$$\begin{aligned} H_L(z) &= a_L(0) \cdot z^{-N} + \sum_{i=1}^{N} a_L(i) \cdot (z^{-N-i} + z^{-N+i}) \\ H_R(z) &= a_R(0) \cdot z^{-N} + \sum_{i=1}^{N} a_R(i) \cdot (z^{-N-i} + z^{-N+i}). \end{aligned} \qquad (6)$$

The stereo residual signals $e_L(k)$ and $e_R(k)$ are the difference between a delayed version of the input signals and the estimate signals $\hat{x}_L(k)$ and $\hat{x}_R(k)$,

$$\begin{aligned} e_L(k) &= x_L(k-N) - a_L(0) \cdot x_M(k-N) - \\ &\quad \sum_{i=1}^{N} a_L(i) \cdot (x_M(k-N-i) + x_M(k-N+i)) \\ e_R(k) &= x_R(k-N) - a_R(0) \cdot x_M(k-N) - \\ &\quad \sum_{i=1}^{N} a_R(i) \cdot (x_M(k-N-i) + x_M(k-N+i)). \end{aligned} \qquad (7)$$

Delaying the input signals is required to compensate the delay introduced by the linear phase filters. For a reconstruction of the stereo signal in the decoder, in addition to the mono signal $x_M(k)$, the two sets of $(N+1)$ stereo prediction coefficients $a_L(i)$ and $a_R(i)$ and the residual signals $e_L(k)$ and $e_R(k)$ are quantized and transmitted. To account for this, in Figure 2, the blocks $Q_{e,R}$, $Q_{H,R}$ for the right channel, $Q_{e,L}$, $Q_{H,L}$ for the left, and Q_M for the mono channel are depicted.

3.1 Optimal Filter Coefficients

For the calculation of the optimal stereo prediction filter coefficients $a_L(i)$ and $a_R(i)$, it is assumed that the signals $x_L(k)$ and $x_R(k)$ are stationary. At first only the right channel is considered.

The target of the optimization procedure is to minimize the expectation of the squared residual signal $e_R(k)$:

$$E\{e_R^2(k)\} \rightarrow \min \qquad (8)$$

The substitution

$$a'_R(i) = \begin{cases} \frac{1}{2} \cdot a_R(i) & \text{for } i = 0 \\ a_R(i) & \text{for } i > 0 \end{cases} \qquad (9)$$

is introduced for the following calculations. With equation (7) and setting its partial derivatives with respect to all $a_R(i)'$ zero yields the following equation:

$$\mathbf{X}_M \cdot \mathbf{a}'_R = \mathbf{X}_{R,M}. \qquad (10)$$

The vector

$$\mathbf{a}'_R = [a'_R(0) \quad a'_R(1) \quad \cdots \quad a'_R(N)]^T \qquad (11)$$

contains the desired filter coefficients. The matrix

$$\mathbf{X}_M = \begin{bmatrix} X_M(0,0) & \cdots & X_M(0,N) \\ \cdots & X_M(j,l) & \cdots \\ X_M(N,0) & \cdots & X_M(N, 2 \cdot N) \end{bmatrix} \qquad (12)$$

is composed of the autocorrelation function values related to the mono signal $x_M(k)$,

$$X_M(j,l) = \varphi_{x_M,x_M}(\mid l-j \mid) + \varphi_{x_M,x_M}(\mid l+j \mid) \qquad (13)$$

with the index l and j to address columns and rows respectively. The vector $\mathbf{X}_{R,M}$ consists of the cross correlation function values,

$$\mathbf{X}_{R,M} = \begin{bmatrix} (\frac{\varphi_{x_R,x_M}(0)+\varphi_{x_R,x_M}(-0)}{2}) \\ (\frac{\varphi_{x_R,x_M}(1)+\varphi_{x_R,x_M}(-1)}{2}) \\ \cdots \\ (\frac{\varphi_{x_R,x_M}(N)+\varphi_{x_R,x_M}(-N)}{2}) \end{bmatrix}. \qquad (14)$$

The optimal filter coefficients $\mathbf{a}'_R$ are hence

$$\mathbf{a}'_R = (\mathbf{X}_M)^{-1} \cdot \mathbf{X}_{R,M} \qquad (15)$$

for the right channel signal which can be efficiently computed by means of the Levinson algorithm. The filter coefficients for the left channel are determined in analogy to (10)-(15) as

$$\mathbf{a}'_L = (\mathbf{X}_M)^{-1} \cdot \mathbf{X}_{L,M}. \qquad (16)$$

3.2 Modification of the New Approach

With the equations to determine the optimal filter coefficients and the relation

$$\varphi_{x_R,x_M}(i) + \varphi_{x_L,x_M}(i) = 2 \cdot \varphi_{x_M,x_M}(i), \tag{17}$$

it can be shown that

$$\begin{aligned} \mathbf{a}'_R + \mathbf{a}'_L &= (\mathbf{X}_M)^{-1} \cdot (\mathbf{X}_{R,M} + \mathbf{X}_{L,M}) \\ &= [1 \quad 0 \quad \cdots \quad 0]^T. \end{aligned} \tag{18}$$

Accordingly, there is a very simple relation between the coefficients for the left and the right channel. In analogy to this, with (7) and (18), a simple relation can be derived for the residual signals for left and right channel as well,

$$e_L(k) + e_R(k) = 0 \; \forall \, k. \tag{19}$$

Considering this result, Figure 2 can be transformed into the diagram shown in Figure 3. As a result, only the filter coefficients $\mathbf{a}_R$, the residual signal $e_R(k)$ and the mono signal $x_M(k)$ must be transmitted to reproduce the left and the right channel signal in the decoder which reduces the required overall bit rate.

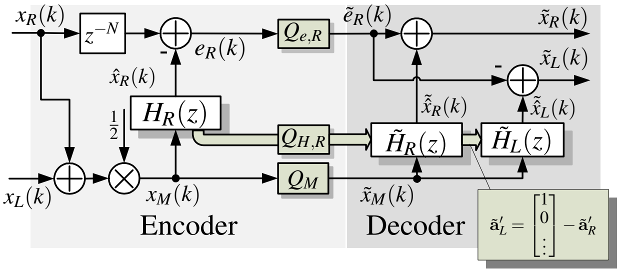


Figure 3: Simplification of the new approach.

3.3 Special Operation Conditions

In the presence of a stereo signal where both channel signals are identical, $x_L(k) = x_R(k)$, the optimal filter coefficients are

$$\mathbf{a}_R = \mathbf{a}_L = [1 \quad 0 \quad \cdots 0]^T \tag{20}$$

so that the residual signal becomes

$$\begin{aligned} e_R(k) &= x_R(k-N) - \frac{x_L(k-N) + x_R(k-N)}{2} \qquad (21) \\ &= \frac{x_R(k-N) - x_L(k-N)}{2} = 0. \qquad (22) \end{aligned}$$

In this case, the new system behaves identically to M/S joint-stereo coding with the side channel signal identical to the stereo residual signal.

In the presence of a signal with a dominant signal in one channel only, e.g. $x_R(k) = 0$, $x_L(k) \neq 0$ the resulting filter coefficients are

$$\mathbf{a}_R = \mathbf{0} \text{ and } \mathbf{a}_L = [2 \quad 0 \quad \cdots 0]^T \tag{23}$$

The residual signal becomes $e_R(k) = e_L(k) = 0$ and the system behaves identically to L/R joint stereo coding with the side channel signal identical to the stereo residual signal. The new approach is hence a generalization of M/S and L/R joint-stereo coding.

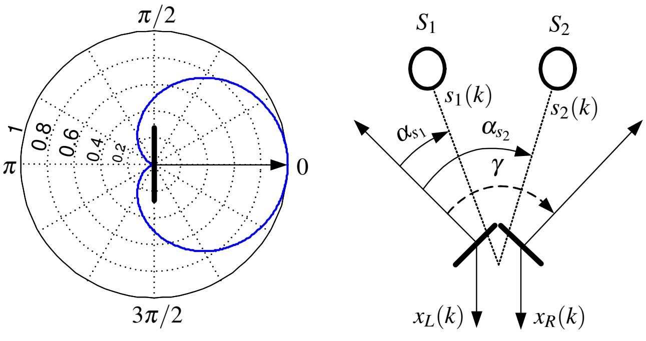
4 Evaluation and Analysis

In the following, the coding performance of the new approach shall be compared to that of conventional M/S joint-stereo coding. The comparison is based on stereo signals which are the output of a stereo signal production model which will at first be described briefly. Referring to the equations in Section 3.1, from the produced stereo signals, the stereo prediction filter coefficients $\mathbf{a}_L$ and $\mathbf{a}_R$ are computed. Given these coefficients, the performance of both approaches can be determined in terms of signal-to-noise ratios (SNR). The target of the system analysis is to compute the SNR related to the reconstruction of the left and the right channel signals, $x_L(k)$ and $x_R(k)$ respectively, in the decoder,

$$SNR_{L/R} = \frac{E\{(x_{L/R}(k))^2\}}{E\{(x_{L/R}(k) - \tilde{x}_{L/R}(k))^2\}}, \tag{24}$$

with L/R representing either the left or the right channel. The measured SNR values will be determined for different overall encoding bit rates per sample, referred to as R_A. Due to the limitation of space in this paper, the theory to calculate the SNR will not be described in detail. Instead, results for two example scenarios will be presented.

In the stereo signal production model, $x_L(k)$ and $x_R(k)$ are assumed to be the output of a stereo recording based on two cardioid microphones [9]. Each cardioid microphone has a directional pickup pattern that is roughly heart-shaped when viewed from above as depicted in Figure 4 a). In order to record stereo signals, the microphones are arranged as depicted in Figure 4 b) with the angle $\gamma = \pi/2$.



a) Directional pattern

b) Stereo recording setup

Figure 4: Directional pattern of cardioid microphone (a) and stereo recording setup (b).

By specifying the angles $0 \leq \alpha_{s_1} \leq \pi/2$ and $0 \leq \alpha_{s_2} \leq \pi/2$, the two stationary, independent signal sources S_1 and S_2 are positioned in the same horizontal plane as the two microphones. The microphones are furthermore assumed to be very close together so that the signal delay between left and right channel is negligible. S_1 and S_2 are modeled as auto-regressive (AR) processes with zero mean and a constant set of AR filter coefficients each. The AR filter coefficients have been determined by analyzing short segments of a real audio sample to well approximate realistic signals. To generate the signals $x_L(k)$ and $x_R(k)$, in the first step, $s_1(k)$ and $s_2(k)$ related to the sources S_1 and S_2 respectively are produced. In the second step, $x_L(k)$ and $x_R(k)$ are computed as a weighted sum of $s_1(k)$ and $s_2(k)$ depending on the angles α_{s_1} and α_{s_2} according to the directivity characteristics of both microphones.

For the determination of SNR_L and SNR_R, the following assumptions are made considering the impact of the quantizers in Figure 3:

- The stereo linear prediction coefficients can be transmitted to the decoder without introducing any quantization error: $\tilde{H}_R(z) \approx H_R(z)$.
- The quantizers for the mono and the prediction error signal, Q_M and $Q_{e,R}$ respectively, are assumed to lead to an SNR that follows the 6-dB-per-bit rule with the bit rates R_M and $R_{e,R}$ per sample respectively.
- A bit allocation prescribes how to decompose the overall bit rate R_A into the bit rates R_M allocated for the quantizer Q_M and $R_{e,R}$ for $Q_{e,R}$ with $R_A = R_M + R_{e,R}$. The bit allocation follows the approach to maximize the sum of the logarithmic SNR in both channels and shall not be discussed in detail here.

M/S joint-stereo coding is realized by setting the coefficients according to equation (20).

5 Results

Two example setups have been investigated which are listed in Table 1. In the case of Setup I, the sources S_1 and S_2 are close to the center position and $s_1(k)$ and $s_2(k)$ have equal signal power. Due to the strong center component, both, the new approach and M/S joint-stereo coding lead to values for SNR_L and SNR_R approximately 6 dB higher than those achieved by coding the left and the right channel signals independently given the same overall bit rate (for bit rates higher than 3 bits per sample).

Setup	α_{s_1}	α_{s_2}	Signal Level
I	$\pi/6$	$\pi/3$	$E\{(s_1(k))^2\} \approx E\{(s_2(k))^2\}$
II	0	$\pi/2$	$10 \cdot \log_{10}(\frac{E\{(s_2(k))^2\}}{E\{(s_1(k))^2\}}) \approx 20\text{dB}$

Table 1: Setups for system evaluation.

The results for Setup II are shown in Figure 5 for different bit rates R_A. The values of SNR_L and SNR_R for the new approach are significantly higher compared to those for M/S coding which shows that the new approach outperforms M/S joint-stereo coding. In the new approach and M/S coding, the SNR for the left channel is lower than that for the right channel. The reason for this is that, due to the unsymmetric setup, the signal power of $x_L(k)$ is higher than that of $x_R(k)$. A very low SNR in one channel leads to annoying audible artifacts on one ear in the case of M/S coding. In the new approach, the SNR unbalance between left and right channel can be partly compensated. Also, the stereo linear prediction filter leads to a frequency selective allocation of the quantization noise related to quantizer Q_M to the left and the right channel. This is not the case in M/S coding. Therefore the new approach yields a stereo noise masking that also leads to a higher perceived audio quality.

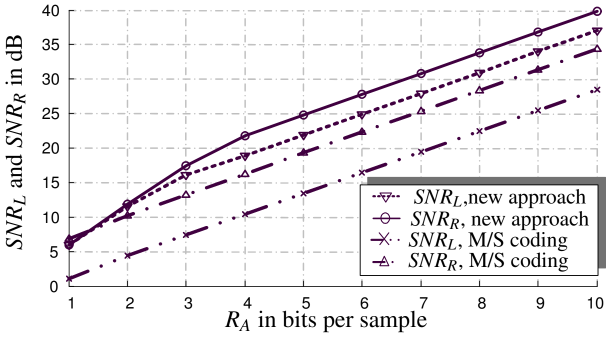


Figure 5: SNR for left and right channel signal, new approach versus M/S joint-stereo coding.

6 Conclusion

In this contribution, a new approach for joint-stereo coding is proposed. The new approach is based on linear prediction techniques to predict the left and the right channel signal from the corresponding mono downmix signal. Since it operates in the time domain a low algorithmic delay can be achieved. In the first part of the paper, equations to calculate the optimal stereo linear prediction coefficients were derived. It was shown that a relation between the optimal stereo prediction coefficients for the left and the right channel exists that can be exploited to reduce the required coding bit rate without any loss. It finally turns out that the new system is a generalization of M/S and L/R joint-stereo coding.

The last part of the paper is a comparison of the coding performance of the new approach and conventional M/S joint-stereo coding. For this purpose, a stereo signal production model was defined making realistic recording assumptions. As result, it was shown that the new approach significantly outperforms M/S joint-stereo coding in terms of calculated SNRs for the left and right channel. In addition to that, a frequency selective allocation of quantization noise to the left and the right channel leads to a stereo noise masking and hence an improved perceived stereo audio quality.

References

[1] A. Biswas. *Advances in Perceptual Stereo Audio Coding Using Linear Prediction Techniques*. PhD thesis, Technische Universiteit Eindhoven, 2007.

[2] J. Breebaart and C. Faller. *Spatial Audio Processing*. John Wiley, 2007.

[3] E.Torick and T.Keller. Improving the signal to noise ratio and coverage of FM stereo broadcasts. *AES Journal*, 33(12), dec 1985.

[4] H. Fuchs. Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction. *IEEE Workshop on Applications of Signal Processing to Audio and Acosutics*, 1993.

[5] J. Herre, K. Brandenburg, and D. Lederer. Intensity Stereo Coding. *AES 96th Conv.*, pages 1–10, 1994.

[6] http://www.answers.com/topic/fm broadcasting. FM broadcasting, 2007.

[7] J.D. Johnston and A.J. Ferreira. Sum-Difference Stereo transform Coding. *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing 1992, San Francisco, USA*, 1992.

[8] T. Liebchen. Lossless Audio Coding Using Adaptive Multichannel Prediction. *113th Conv. of the Audio Eng. Soc.(AES), Los Angeles, USA*, 2002.

[9] E. Zwicker and M. Zollner. *Elektroakustik*. Springer, Berlin, 1998.