DELAYLESS SOFT-DECISION DECODING OF HIGH-QUALITY AUDIO WITH ADAPTIVELY SHAPED PRIORS

Florian Pflug and Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, D–38106 Braunschweig, Germany
phone: +49 531 391-2476, fax: +49 531 391-8218, e-mail: {f.pflug,t.fingscheidt}@tu-bs.de

ABSTRACT

For high-quality digital audio transmitted over error-prone short-range wireless channels robust source decoding with low decoding delay is desired. However, many previous approaches from the field of audio error concealment are solely intuitively motivated or introduce algorithmic delay. In contrast, this paper deals with a Bayesian framework for delayless full-band soft-decision error concealment for quantized but uncompressed audio utilizing only residual redundancy in the audio signal and channel reliability information. In principle, it can be applied to any channel providing reliability information. As a novelty, we employ multiple adaptively shaped prior probability distributions in the decoding process. These are utilized in conjunction with the autocorrelation method or the normalized least-mean-square (NLMS) algorithm in order to compute prediction probabilities within the Bayesian framework. Experiments carried out on representative audio data transmitted over additive white Gaussian noise (AWGN) channels show significant enhancements in audio quality.

1. INTRODUCTION

Robust source decoding algorithms are required in digital communications wherever signals are transmitted over error-prone channels. This is due to the fact that even single bit errors after a possible error correction scheme can lead to annoying distortions. For example, a flipped most significant bit of a pulse-code modulated (PCM) sample leads to a loud click in the corresponding audio signal. In addition, advanced channel coding algorithms like turbo codes or low-density parity check (LDPC) codes introduce a waterfall effect, i. e., above a certain signal-to-noise ratio (SNR) residual bit errors occur only very rarely, whereas below this threshold error rates rise tremendously. Therefore, it is reasonable to employ robust error concealment techniques in order to introduce a graceful degradation in signal quality.

Numerous previous approaches exist for efficient error concealment of source-coded signals (see, e. g., [1–6]). However, these signals are either narrowband speech signals coded with a low bit rate or signals coded with considerable delays due to efficient source coding algorithms (e. g., MPEG-1 Audio Layer 2 [7] or MPEG-4 Part 3 AAC [8]). In contrast, upcoming digital wireless audio transmission systems like WirelessHD [9], Wireless Home Digital Interface (WHDI) [10] or digital wireless head- and microphones require *very low* end-to-end delays while transmitting CD-quality audio signals, i. e., signals sampled at a minimum sampling rate of 44.1 kHz and quantized with 16 bits per sample. For professional grade and audiophile applications even sampling rates of 96 or 192 kHz and quantization with 24 bits per sample are common. Due to delay and/or complexity restrictions, the signals are often transmitted without computationally expensive source coding or even in uncompressed form. As a result, the complexity for robust decoding would increase greatly if known error concealment approaches such as those presented in [3–6] were applied.

Error concealment strategies like denoising, declicking and decrackling of distorted audio signals have been widely investigated (see, e. g., [11–15]). However, these algorithms process hard-decided signals and do not benefit from a bit-wise reliability information obtained during demodulation or channel decoding. In contrast, in [2] a Bayesian approach for robust speech decoding exploiting bit-wise reliability information together with residual redundancy in a speech signal has been presented. We have shown in [16] that a modified approach also yields significant performance gains in delayless soft-decision decoding of high-quality audio signals. Precisely, we have presented two novel approaches for the computation of prediction probabilities for audio samples within a Bayesian framework, the autocorrelation/Levinson-Durbin method and the normalized least-mean-square (NLMS) approach.

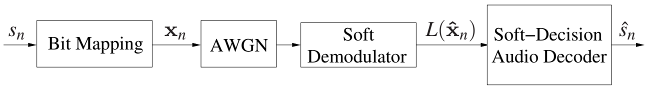

Figure 1: High-level block diagram of the simulation setup

In this contribution, we introduce the employment of adaptively shaped, amplitude-dependent pre-trained *a priori* probability distributions for efficient error concealment of high-quality audio. This means that multiple prediction error probability density functions (PDFs) dependent on the amplitude range of a predicted audio sample are generated in a training process. These PDFs are employed in the decoding process, taking the different signal statistics of quiet and loud parts of an audio signal into account. As our investigations are focused on short-range line-of-sight scenarios, we employ an additive white Gaussian noise (AWGN) channel. In addition, we do not employ any channel coding. The extension of our approach to more complex channels yielding reliability information is straightforward.

The organization of the paper is as follows: Section 2 describes the general steps of soft-decision audio decoding. Our new approach to compute prediction probabilities with adaptively shaped *a priori* knowledge is presented in Section 3. Simulation results are discussed in Section 4. Finally, Section 5 draws conclusions with regard to the achieved results.

2. SOFT-DECISION AUDIO DECODING

The basic framework for soft-decision audio decoding is based on the approach presented in [4]. The simulation setup is depicted in Fig. 1 and starts with audio samples $s_n \in \{-1, -1+\Delta, \dots, 1-\Delta\}$ quantized in M bit resolution, with $n \in \{0, 1, \dots\}$ denoting the sample index and $\Delta = 2^{-M+1}$ being the quantization step size. After a mapping to a bit combination $\mathbf{x}_n = (x_n(0), x_n(1), \dots, x_n(m), \dots, x_n(M-1))$ with bipolar bits $x_n(m) \in \{-1, 1\}$, the signals are transmitted over an AWGN channel. A soft demodulator yields log-likelihood ratios (LLRs)

$$L(\hat{x}_n(m)) = \ln \frac{\mathrm{P}(\hat{x}_n(m)|x_n(m) = +1)}{\mathrm{P}(\hat{x}_n(m)|x_n(m) = -1)} \tag{1}$$

for each transmitted bit $\hat{x}_n(m)$. These LLRs serve to compute bit error probabilities $p_{e,n}(m) = 1/(1 + \exp(|L(\hat{x}_n(m))|))$ for each hard-decided bit $\hat{x}_n(m) = \mathrm{sign}(L(\hat{x}_n(m)))$ on the receiver side. Assuming

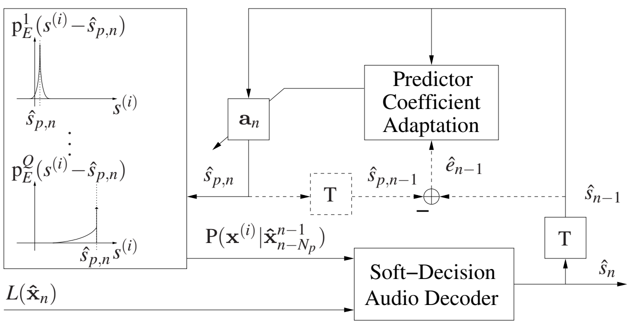

Figure 2: Block diagram for the computation of prediction probabilities $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1})$ with a number of Q adaptively shaped PDFs. Dashed lines indicate paths only necessary for the normalized least-mean-square (NLMS) approach.

a memoryless channel, this reliability information is used for the computation of transition probabilities according to

$$P(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)}) = \prod_{m=0}^{M-1} P(\hat{x}_n(m)|x^{(i)}(m)), \tag{2}$$

with

$$P(\hat{x}_n(m)|x^{(i)}(m)) = \begin{cases} p_{e,n}(m) & \text{if } \hat{x}_n(m) \neq x^{(i)}(m), \\ 1 - p_{e,n}(m) & \text{else.} \end{cases} \tag{3}$$

This describes the probability for a received hard-decided bit combination $\hat{\mathbf{x}}_n$ given that the bit combination $\mathbf{x}^{(i)}$ was transmitted, with $i \in \{0, 1, \ldots, 2^M - 1\}$.

For an estimation $\hat{s}_n$ of the transmitted audio sample s_n the minimum mean-square error (MMSE) $E\{(s_n - \hat{s}_n)^2\}$ is used as an optimization criterion, with $E\{\cdot\}$ denoting the expectation value. The corresponding MMSE estimation rule can be written as

$$\hat{s}_n = \sum_{i=0}^{2^M-1} s^{(i)} \cdot P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}), \tag{4}$$

with $(\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}) = (\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_{n-1}, \ldots, \hat{\mathbf{x}}_0)$ being the complete history of received bit combinations and $s^{(i)} \in \{-1, -1+\Delta, \ldots, 1-\Delta\}$, $\Delta = 2^{-M+1}$, denoting all possibly transmitted samples in M bit resolution, represented by bit combination $\mathbf{x}^{(i)}$. An advantage of the MMSE estimator is the inherent muting effect. It results from the *a posteriori* probabilities becoming $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}) = P(\mathbf{x}^{(i)})$ for the worst case channel with $p_{e,n}(m) = 0.5$ and the assumption that audio signals have a zero mean. As a result, estimated sample values are attenuated leading to a graceful degradation of audio quality for very bad channel states.

A posteriori probabilities $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1})$ describe the probabilities for each possibly transmitted bit combination $\mathbf{x}^{(i)}$ given the received bit combinations $\hat{\mathbf{x}}_n$, $\hat{\mathbf{x}}_0^{n-1}$. As only present and past bit combinations are regarded, no algorithmic delay is introduced in this step. Assuming a memoryless channel, *a posteriori* probabilities can be determined according to

$$P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}) = \frac{1}{C} \cdot P(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)}) \cdot P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}), \tag{5}$$

with $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1})$ denoting prediction probabilities and C being a constant, such that $\sum_{i=0}^{2^M-1} P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}) = 1$.

Prediction probabilities comprise receiver-side *a priori* knowledge about a possibly transmitted sample value $\mathbf{x}^{(i)}$ before $\hat{\mathbf{x}}_n$ has been received. Without any *a priori* knowledge all $\mathbf{x}^{(i)}$ are uniformly distributed and the prediction probability reduces to $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) = 2^{-M}$. This is referred to as NAK (no *a priori* knowledge) in the following. If the 0^{th} order distribution of audio sample values is known, e. g., by measurements in a training process, prediction probabilities can be written as $P(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) = P(\mathbf{x}_n = \mathbf{x}^{(i)})$. As a consequence, for digital audio quantized with 16 bits per sample, a memory size of $2^M = 2^{16}$ words is required for the decoder. This is denoted with AK0 in the following.

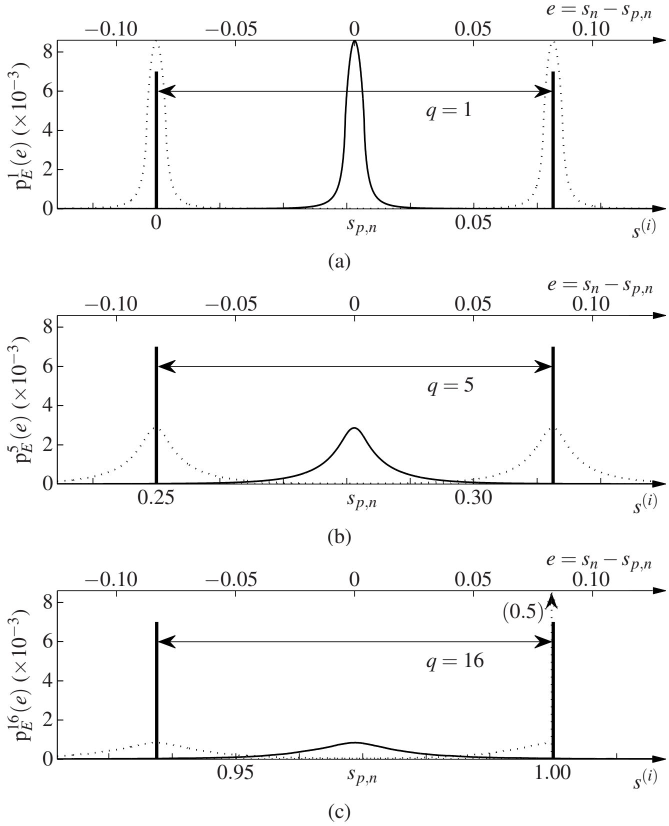

Figure 3: Examples of probability density functions $p_E^q(e = s_n - s_{p,n})$ with $Q = 16$ and (a) $q = 1$, (b) $q = 5$ and (c) $q = 16$. Solid vertical bars mark the limits of $s_{p,n}$ within an interval q, dashed PDFs illustrate the maximum possible shifted positions of the corresponding PDFs.

3. APPROACH TO PREDICTION PROBABILITIES BY LINEAR PREDICTION

In order to perform efficient error concealment for audio it is obvious to include information from preceding estimated audio samples into the estimation process. This can be accomplished by employing linear prediction schemes in the decoder [2, 16].

The audio signal is predicted according to Fig. 2 by a linear combination of N_p preceding estimated audio samples $\hat{\mathbf{s}}_{n-N_p}^{n-1} = (\hat{s}_{n-1}, \hat{s}_{n-2}, \ldots, \hat{s}_{n-N_p})^T$, with N_p denoting the prediction order and $(\cdot)^T$ being a transposed vector. The predicted sample is obtained by

$$\hat{s}_{p,n} = \mathbf{a}_n^T \cdot \hat{\mathbf{s}}_{n-N_p}^{n-1}, \tag{6}$$

with $\mathbf{a}_n = (a_n(1), \ldots, a_n(N_p))^T$ denoting the N_p predictor coefficients. $\mathbf{a}_n$ is adapted for each sample index in order to minimize the mean-square error $E\{\hat{e}_n^2|\hat{\mathbf{s}}_{n-N_p}^{n-1}\}$, with $\hat{e}_n = \hat{s}_n - \hat{s}_{p,n}$ denoting the prediction error. Two approaches for this are described briefly in the following.

3.1 Predictor Coefficient Adaptation

3.1.1 Autocorrelation Approach

The autocorrelation (ACOR) method is a well-known approach for the computation of predictor coefficients [17]. First, L previously estimated audio samples $\hat{\mathbf{s}}_{n-L}^{n-1}$ are windowed according to $\mathbf{W}\cdot\hat{\mathbf{s}}_{n-L}^{n-1}$, with $\mathbf{W} = 0.54\cdot\mathbf{I} - 0.46\cdot\mathrm{diag}\{\cos(\frac{2\pi\cdot L}{2L}),\ldots,\cos(\frac{2\pi\cdot 1}{2L})\}$ denoting a left-half Hamming window, $\mathbf{I}$ being the identity matrix and $\mathrm{diag}\{\cdot\}$ denoting a diagonal matrix with its diagonal elements defined by its argument. These windowed samples are then used for the computation of N_p+1 autocorrelation coefficients. Finally, the predictor coefficients $\mathbf{a}_n$ are determined by the Levinson-Durbin algorithm applied to a Toeplitz matrix of the autocorrelation coefficients [18]. The predictor coefficients are updated according to this scheme for each sample index.

3.1.2 Normalized Least-Mean-Square Approach

For adaptive filtering and source coding of audio signals the normalized least-mean-square (NLMS) algorithm is widely used [18–20]. The adaptation rule for the predictor coefficients can be written as

$$\mathbf{a}_n = \mathbf{a}_{n-1} + \frac{\hat{e}_{n-1}}{1+\lambda\cdot||\hat{\mathbf{s}}_{n-N_p}^{n-1}||^2}\cdot\hat{\mathbf{s}}_{n-N_p}^{n-1}\,, \tag{7}$$

with λ denoting a tuning parameter controlling the convergence rate [19]. The NLMS algorithm uses the MMSE between the filter output $\hat{s}_{p,n}$ and the desired response $\hat{s}_n$ as an optimization criterion. Due to the normalization with respect to the squared Euclidean norm $||\hat{\mathbf{s}}_{n-N_p}^{n-1}||^2$ the coefficient update is only loosely dependent on the energy of $\hat{\mathbf{s}}_{n-N_p}^{n-1}$. This is especially important for high-dynamic audio signals. The variables are initialized at sample index $n=0$ with $\mathbf{a}_{-1} = (1/N_p,\ldots,1/N_p)^T$ and $\hat{\mathbf{s}}_{-N_p}^{-1} = (0,\ldots,0)^T$ [19].

3.2 New Adaptively Shaped Priors

The predicted sample $\hat{s}_{p,n}$ serves to select and shift a pre-trained and stored prediction error probability density function (PDF), yielding the prediction probabilities $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1})$ of the Bayesian framework in (5).

First, the statistics of the prediction error $e_n = s_n - s_{p,n}$ is measured in a training process. Depending on the magnitude of the corresponding predicted sample $s_{p,n}$, each e_n value is utilized for the training of one of Q amplitude-dependent prior distributions $\mathrm{p}_E^q(e)$, with $q\in\{1,\ldots,Q\}$. The actual PDF is selected by $q=\lfloor Q\cdot|s_{p,n}|+1\rfloor$. Only the magnitude of $s_{p,n}$ is taken into account in this step because audio signals are assumed to be symmetrical.

As we are interested in computing the prediction probabilities $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) \approx \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1})$ in (5), we note that $\mathbf{x}^{(i)}$ and $\hat{\mathbf{x}}_{n-N_p}^{n-1}$ can be directly mapped to signals $s^{(i)}$ and $\hat{\mathbf{s}}_{n-N_p}^{n-1}$, allowing us to write

$$\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1}) = \mathrm{P}(s^{(i)}|\hat{\mathbf{s}}_{n-N_p}^{n-1})\,. \tag{8}$$

The sole influence of the past estimated samples $\hat{\mathbf{s}}_{n-N_p}^{n-1}$ on the MMSE estimate $\hat{s}_n$ can be summarized in the predicted value $\hat{s}_{p,n} = E\{\hat{s}_n|\hat{\mathbf{s}}_{n-N_p}^{n-1}\}$. Therefore we can write

$$\mathrm{P}(s^{(i)}|\hat{\mathbf{s}}_{n-N_p}^{n-1}) = \mathrm{P}(s^{(i)}|\hat{s}_{p,n})\,, \tag{9}$$

with $\mathrm{P}(s^{(i)}|\hat{s}_{p,n})$ being the PDF of the prediction error. Using a number of Q different prediction error PDFs dependent on the amplitude of the predicted sample, generated as described above and stored as priors, we finally can write

$$\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1}) = \mathrm{P}(s^{(i)}|\hat{s}_{p,n}) \tag{10}$$

$$= \mathrm{p}_E^{\hat{q}}(s^{(i)}-\hat{s}_{p,n}) \quad \text{with} \quad \hat{q} = \lfloor Q\cdot|\hat{s}_{p,n}|+1\rfloor\,. \tag{11}$$

The aim of this novel approach with adaptively shaped *a priori* PDFs for soft-decision audio decoding is to take advantage of the different signal characteristics (predictability) of audio signals within a certain amplitude range. For example, the amplitudes of adjacent samples are all expected to be rather small for quiet parts of an audio signal. As a result, linear prediction yields small prediction errors and the shape of the corresponding PDF is narrow (see Fig. 3(a)). Conversely, in loud parts the audio signal is more likely highly non-stationary, i. e., the amplitudes of neighbored samples are more likely large but still low-amplitude samples occur. In these cases, dependent on the adaptation rate of the predictor coefficient update algorithm, prediction errors become larger. Consequently, the corresponding PDF has a broader shape (see Fig. 3(b)). Additionally, in Fig. 3 the maximum shifted positions of the PDFs for the given intervals q are shown with dashed lines, solid lines denote the central PDF position.

Furthermore, very high audio sample magnitudes greater than 1 simply cannot occur and have a probability of zero, leading to an asymmetrical shifted PDF (e. g., dashed PDF on the right-hand side in Fig. 3(c)). In this case, the values of the clipped PDF tail are summed up and inserted as a corresponding weighted unit impulse (illustrated by the dashed arrow in Fig. 3(c)).

4. SIMULATIONS

4.1 Experimental Setup

We transmitted 13 monaural audio signals having a total length of 96 s from various sources (passages from classical pieces and a motion picture sound track with music and effects, with different instruments like organs, brass instruments, strings, percussions, pianos and synthesizers) over an AWGN channel with E_b/N_0 ratios ranging from 0–12 dB, as seen in Fig. 1. *A priori* probabilities $\mathrm{P}(\mathbf{x}^{(i)})$ and $Q=16$ prior distributions $\mathrm{p}_E^q(e)$ have been gathered from training material comprising 15 musical pieces used only for training (pieces of classical music, electronic music and a motion picture soundtrack including music, speech and effects) with a total length of 81 min. All audio signals have been normalized to –26 dBFS (dB relative to full-scale), are sampled with 48 kHz and are linear pulse-code modulated (PCM) with 16 bits per sample. In order to measure the decoding performance, we use the global signal-to-noise ratio (SNR)

$$\mathrm{SNR}_{\mathrm{global}} = 10\log_{10}\frac{E\{\tilde{s}^2\}}{E\{(\hat{s}-\tilde{s})^2\}}\,, \tag{12}$$

and the segmental SNR

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{10}{K}\sum_{k=0}^{K-1}\log_{10}\frac{\sum_{n=0}^{N-1}\tilde{s}_{n+k\cdot N}^2}{\sum_{n=0}^{N-1}(\hat{s}_{n+k\cdot N}-\tilde{s}_{n+k\cdot N})^2}\,, \tag{13}$$

with K being the number of frames and $N=480$ (i. e., 10 ms) denoting the frame length for evaluation. In the field of error concealment schemes both measures have been proven to be notably useful [2, 4]. $\mathrm{SNR}_{\mathrm{seg}}$ serves as an indicator for permanent but low-amplitude artifacts like crackling; $\mathrm{SNR}_{\mathrm{global}}$ reflects rare transient artifacts like sporadic loud clicks. *Only* for the SNR measurements reference signal $\tilde{s}$ is available in 24 bit resolution. Assuming a perfectly Laplacian-distributed signal at –26 dBFS and fine uniform quantization ($E\{(\hat{s}-\tilde{s})^2\}\approx\Delta^2/12$), a maximum global SNR of approximately 77 dB can be achieved.

In the preparatory stage of our experiments we made several numerical investigations regarding well-suited window sizes L, prediction orders N_p and convergence parameters λ. Our analyses showed that $\lambda=20$ for the NLMS, $L=960$ for the ACOR approach and $N_p=10$ for both approaches are reasonable choices regarding signal-to-noise ratios. Furthermore, it turned out that for the worst investigated channel state it is more sensible to utilize the previously estimated sample value $\hat{s}_{n-1}$ instead of the prediction value

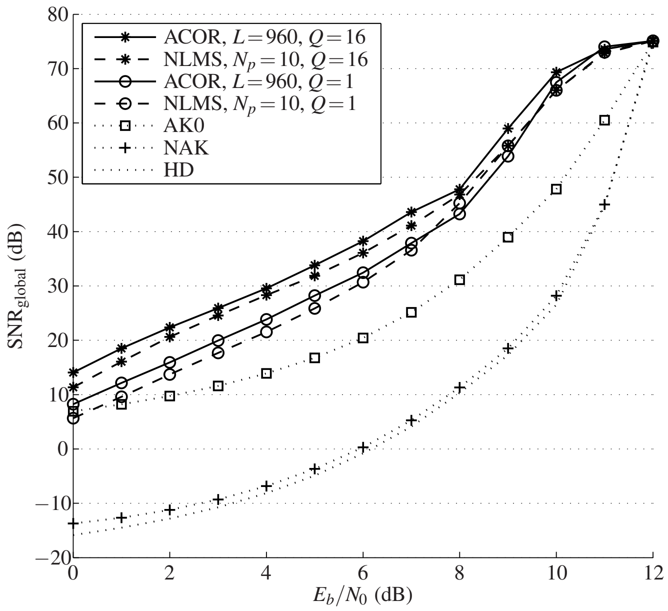

Figure 4: Global SNR performance with soft-decision decoding using the autocorrelation (ACOR) and NLMS approaches with $Q = 16$ and $Q = 1$ priors, 0^{th} order *a priori* knowledge (AK0), no *a priori* knowledge (NAK) and hard-decision decoding (HD).

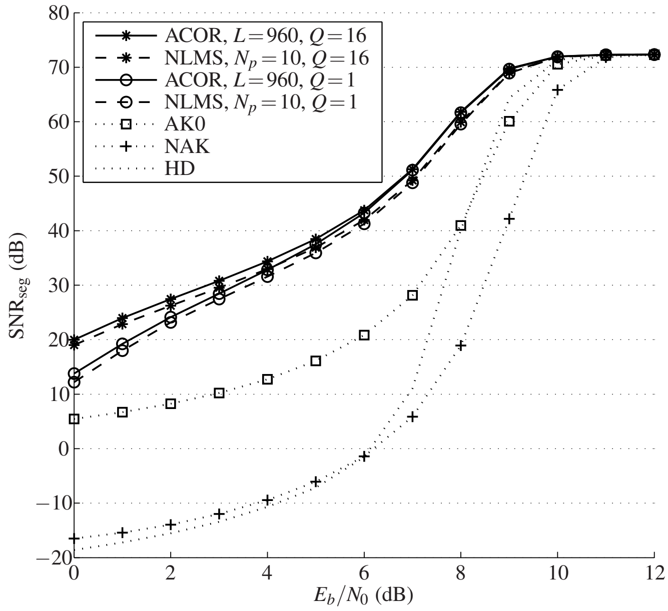

Figure 5: Segmental SNR performance with soft-decision decoding using the autocorrelation (ACOR) and NLMS approaches with $Q = 16$ and $Q = 1$ priors, 0^{th} order *a priori* knowledge (AK0), no *a priori* knowledge (NAK) and hard-decision decoding (HD).

$\hat{s}_{p,n}$ for the *selection* of the prior distribution. For all other simulated channel states, no differences in either $\mathrm{SNR_{global}}$ or $\mathrm{SNR_{seg}}$ were measurable. Therefore, we used $\hat{q} = \lfloor Q \cdot |\hat{s}_{n-1}| + 1 \rfloor$ throughout *all* E_b/N_0 conditions in our experiments. Note that the prediction error PDF *shift* $\mathrm{p}_E^{\hat{q}}(s^{(i)} - \hat{s}_{p,n})$ is of course performed with the predicted sample $\hat{s}_{p,n}$.

In the current work, we investigate the parameters $Q \in \{1, 16\}$ for the ACOR and NLMS approaches. This means that either only a single prior distribution ($Q = 1$) or multiple adaptively shaped prior distributions ($Q = 16$) are employed for the computation of prediction probabilities in the estimation process.

4.2 Discussion

The simulation results are depicted in Figs. 4 and 5 for the ACOR and NLMS soft-decoding approaches with $Q \in \{1, 16\}$, for soft-decision decoding with 0^{th} order *a priori* knowledge (AK0) and without *a priori* knowledge (NAK), and for hard-decision decoding (HD). The channel E_b/N_0 ratios are given on the abscissae, the resulting SNRs after error concealment are provided on the ordinates.

Our results show that soft-decision audio decoding with adaptively shaped priors ($Q = 16$) is consistently superior to the corresponding conventional approach with only $Q = 1$ *a priori* probability distribution with virtually no notable increase in complexity. In fact, improvements of up to 2 dB in E_b/N_0 and 6 dB in both $\mathrm{SNR_{global}}$ and $\mathrm{SNR_{seg}}$ can be reached. Compared to HD decoding, even gains of 8.5 dB in E_b/N_0, 40 dB in $\mathrm{SNR_{global}}$ and 45 dB in $\mathrm{SNR_{seg}}$ are achieved. At very high E_b/N_0 ratios all decoding algorithms approach $\mathrm{SNR_{global}} = 75.1$ dB and $\mathrm{SNR_{seg}} = 72.1$ dB, respectively, which corresponds to the signal-to-noise ratios of the error-free transmission.

As shown in Fig. 5, simple HD decoding outperforms the NAK algorithm in $\mathrm{SNR_{seg}}$ performance for E_b/N_0 ratios greater than 5 dB. This is because distortions become less likely for rising E_b/N_0 ratios but the NAK algorithm constantly introduces audible noise due to the effectively uniformly distributed *a priori* probabilities. This shows that exploitation of channel reliability information alone without the utilization of residual redundancy in the audio signal does not lead to satisfying results. Furthermore, for the worst investigated channel state with an E_b/N_0 ratio of 0 dB the AK0 algorithm reaches the $\mathrm{SNR_{global}}$ performance of both ACOR and NLMS approaches with $Q = 1$. This is due to the fact that it is more meaningful to use a correct amplitude distribution $\mathrm{P}(\mathbf{x}^{(i)})$ than exploiting the correlation of unreliably estimated samples. However, the introduction of $Q = 16$ adaptively shaped priors leads to a performance gain compared to AK0 decoding of approximately 7 dB for the ACOR approach and 4 dB for the NLMS algorithm, respectively. Informal listening tests confirm these instrumental results.

Please note that soft-decision decoding with *a priori* knowledge (AK0, NLMS, ACOR) always leads to SNRs greater or equal than 0 dB. This is due to the inherent muting effect of the MMSE estimator (see Sec. 2).

Regarding the results for $\mathrm{SNR_{seg}}$ in Fig. 5, it can be seen that the performance gain for both $Q = 16$ approaches is large compared to the corresponding $Q = 1$ approaches for bad channel qualities and diminishes for rising E_b/N_0 ratios. On the other hand, Fig. 4 reveals that the performance gap in $\mathrm{SNR_{global}}$ between both approaches remains considerable throughout a wide E_b/N_0 range. In fact, the performance gain in $\mathrm{SNR_{global}}$ ranges between 4.5–6 dB for the NLMS approach and is constantly 6 dB for the ACOR approach, respectively, between E_b/N_0 ratios of 0–7 dB. This suggests that the employment of adaptively shaped priors in the estimation process especially supports the concealment of strong distortions.

5. CONCLUSIONS

In this paper we have presented a fully Bayesian approach to soft-decision error concealment of high-quality audio signals. As a novelty, we have employed multiple adaptively shaped, amplitude-dependent prior distributions in the estimation process. The soft-decoding algorithm uses channel reliability information and prediction probabilities gained by linear prediction for the computation of *a posteriori* probabilities. Predictor coefficients needed for linear prediction are updated by either the autocorrelation/Levinson-Durbin method or a normalized least-mean-square (NLMS) approach. For each sample index, a prior distribution required for the Bayesian framework is determined by the currently predicted audio sample. We have investigated the effects of employing multiple prior distributions, each one trained for a specific amplitude range of predicted samples. Thereby, we take account of the different characteristics of audio signals within different amplitude ranges. Simulations with an AWGN channel show that the suggested approach yields improvements of up to 2 dB in E_b/N_0 compared to our previous works without introducing noticeable computational overhead. Our approach can be employed in a wide range of audio applications such as audio streaming in WirelessHD, Wireless Home Digital Interface (WHDI) or in wireless microphones or headphones.

REFERENCES

[1] K. Sayood and J. C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," *IEEE Trans. Commun.*, vol. 39, no. 6, pp. 838–846, Jun. 1991.

[2] T. Fingscheidt and P. Vary, "Robust Speech Decoding: A Universal Approach to Bit Error Concealment," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, vol. 3, Munich, Germany, Apr. 1997, pp. 1667–1670.

[3] M. Adrat, J. Spittka, S. Heinen, and P. Vary, "Error Concealment by Near Optimum MMSE-estimation of Source Codec Parameters," in *Proc. of IEEE Workshop on Speech Coding*, Delavan, WI, U.S.A., Sep. 2000, pp. 84–86.

[4] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 3, pp. 240–251, Mar. 2001.

[5] F. Lahouti and A. K. Khandani, "Soft Reconstruction of Speech in the Presence of Noise and Packet Loss," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 44–56, Jan. 2007.

[6] N. Phamdo and N. Farvardin, "Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels — Applications to Combined Source-Channel Coding," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 186–193, Jan. 1994.

[7] "Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio," ISO/IEC 13818-3, Apr. 1998.

[8] "Information technology – Coding of audio-visual objects – Part 3: Audio," ISO/IEC 14496-3, Sep. 2009.

[9] WirelessHD Consortium, WirelessHD Specification 1.0a, Aug. 2009.

[10] WHDI Special Interest Group, Wireless Home Digital Interface Specification 1.0, Dec. 2009.

[11] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration - A Statistical Model Based Approach*. Berlin, Germany: Springer, 1998.

[12] C. Févotte, B. Torrésani, L. Daudet, and S. J. Godsill, "Sparse Linear Regression With Structured Priors and Application to Denoising of Musical Audio," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 174–185, Jan. 2008.

[13] N. Tatlas, A. Floros, T. Zarouchas, and J. Mourjopoulos, "An Error–Concealment Technique for Wireless Digital Audio Delivery," in *Proc. of 5th Int. Conf. on Commun. Sys., Networks, and Dig. Signal Process.*, Patras, Greece, Jul. 2006, pp. 181–184.

[14] P. K. Ramarapu and R. C. Maher, "Methods for Reducing Audible Artifacts in a Wavelet-Based Broad-Band Denoising System," *J. Audio Eng. Soc.*, vol. 46, no. 3, pp. 178–190, Mar. 1998.

[15] E. V. Harinarayanan, D. Sinha, S. Saeed, and A. Ferreira, "A Novel Automatic Noise Removal Technique for Audio and Speech Signals," in *Preprints of 123rd Audio Eng. Soc. Con.*, New York, NY, U.S.A., Oct. 2007.

[16] F. Pflug and T. Fingscheidt, "Delayless Soft-Decision Decoding of High-Quality Audio Transmitted over AWGN Channels," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 489–492.

[17] J. E. Markel and A. H. Gray, *Linear Prediction of Speech*, ser. Communication and Cybernetics 12. New York, NY, U.S.A.: Springer, 1976.

[18] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ, U.S.A.: Prentice-Hall, 2002.

[19] G. D. T. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual Audio Coding Using Adaptive Pre- and Post-Filters and Lossless Compression," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 379–389, Sep. 2002.

[20] H. Huang, P. Fränti, D. Huang, and S. Rahardja, "Cascaded RLS-LMS Prediction in MPEG-4 Lossless Audio Coding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 554–562, Mar. 2008.