ON THE USE OF EXPLICIT REDUNDANCY FOR DELAYLESS SOFT-DECISION AUDIO DECODING

Florian Pflug and Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, 38106 Braunschweig, Germany
E-mail: {f.pflug,t.fingscheidt}@tu-bs.de

ABSTRACT

Wireless transmission systems for high-quality digital audio signals require a low end-to-end delay and strong robustness against channel distortions. In this work we investigate a Bayesian approach to delayless soft-decision decoding of high-quality audio signals jointly exploiting both implicit redundancy within the audio signal and explicit sample-wise redundancy appended by a channel (block) encoder. Because our approach introduces no algorithmic delay, it can be employed in audio transmission systems that are extremely sensitive to latency like, e. g., wireless digital microphones. Experiments carried out with representative audio signals transmitted over AWGN channels show a significant increase in signal quality.

***Index Terms*—** soft-decision audio decoding, delayless, linear prediction, block code

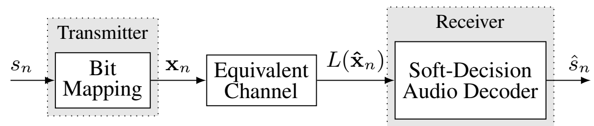


Fig. 1: High-level block diagram of the simulation setup.

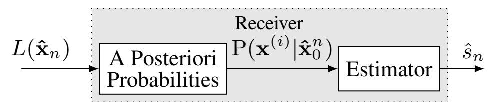


Fig. 2: Block diagram of the soft-decision audio decoder.

1. INTRODUCTION

Wireless transmission systems for high-quality audio signals require strong robustness against channel distortions and a low end-to-end delay. As a result, efficient approaches to error concealment are necessary, otherwise even a single residual bit error could lead to an unacceptable degradation of audio quality for the listener.

Uncompressed high-quality audio signals sampled at $f_s \geq 48\,\text{kHz}$ and quantized with $M \geq 16$ bits per sample exhibit a large amount of residual redundancy. Instead of removing this redundancy by source encoding and later appending it again explicitly by channel coding, this redundancy can be directly exploited for error concealment [1]. As a result, a large number of approaches to robust source decoding and joint source-channel decoding exploiting residual redundancy exist (see, e. g., [2–7]). These approaches work either on source-coded or on narrow-band signals with a low number of possible sample values, allowing them to apply Markov chains in order to exploit statistical properties of the transmitted signal. However, this is not possible for fine-quantized audio signals with high bit rates due to the exponential rise of complexity [8].

In order to increase the robustness of digital transmissions, usually channel coding schemes with corresponding soft-output channel decoders (e. g., [9–13]) are employed. However, these schemes introduce a certain amount of algorithmic delay, e. g., because relatively large block lengths are required for decent error correction performance. In addition, efficient channel codes such as low-density parity-check (LDPC) codes [12] or turbo codes [13] can tremendously increase the complexity of a receiver because of their iterative decoding processes.

An efficient approach to soft-decision decoding of fine-quantized audio signals employing high-order linear prediction has recently been presented in [8, 14]. This approach is based on a Bayesian framework introduced in [5, 15] and allows us to efficiently exploit both the *implicit* redundancy within an audio signal (by a time-variant prior term) and reliability information available from the channel (likelihood term) with low or even no algorithmic delay. In this work we investigate the effects of exploiting additional sample-wise *explicit* redundancy in terms of block codes within the Bayesian framework in the context of prediction-based robust decoding of fine-quantized audio signals. Due to the fact that these block codes are built upon just a single audio sample, the resulting scheme still does not introduce any algorithmic delay.

The paper is organized as follows. In Sec. 2 a brief summary of our soft-decision decoding approach for high-quality audio and the underlying Bayesian framework is provided. Sec. 3 describes the employment of sample-individual block codes and the corresponding delayless decoding approach within the Bayesian framework. The simulation setup and the evaluation of the proposed approach is presented in Sec. 4. Finally, the paper is concluded in Sec. 5.

2. BAYESIAN FRAMEWORK FOR SOFT-DECISION AUDIO DECODING

Our basic simulation setup consisting of a transmitter, an equivalent channel and a receiver is presented in Figs. 1 and 2. In the transmitter, audio samples $s_n \in \{-1, -1+\Delta, \ldots, 1-\Delta\}$ quantized with a resolution of M bits are mapped to bit combinations $\mathbf{x}_n = (x_n(0), x_n(1), \ldots, x_n(m), \ldots, x_n(M-1))$, with $n \in \{0, 1, \ldots\}$ denoting the sample index, $\Delta = 2^{-M+1}$ being the quantization step size and $x_n(m) \in \{0, 1\}$ denoting a single bit. These bits are the input of the equivalent channel, which comprises modulation, the transmission channel, and (soft-output) demodulation (e. g., [4, 16]). The decoder receives log-likelihood ratios (LLRs) $L(\hat{\mathbf{x}}_n)$ for each hard-decided received bit $\hat{x}_n(m) = 0.5 \cdot (1 - \mathrm{sign}(L(\hat{x}_n(m))))$,

978-1-4799-0356-6/13/$31.00

which are utilized to compute bit error probabilities $p_{e,n}(m) = 1/(1+\exp(|L(\hat{x}_n(m))|))$ (e. g., [4, 5, 11]). With the assumption of a memoryless channel, we can now determine the likelihood term of the Bayesian framework, the so-called *transition probabilities*

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

with

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

The term $\mathrm{P}(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)})$ describes the probability for transition from a possibly transmitted bit combination $\mathbf{x}^{(i)}$, with $i \in \{0, 1, \ldots, 2^M - 1\}$, to the received bit combination $\hat{\mathbf{x}}_n$. Thereby, it comprises all available channel reliability information expressed by the LLRs.

2.1. A Posteriori Probabilities

In order to estimate the transmitted audio sample, *a posteriori* probabilities $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^n)$ for each $\mathbf{x}^{(i)}$ with respect to the complete receive history $\hat{\mathbf{x}}_0^n = (\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n)$ have to be determined. These probabilities can be split up by applying Bayes' theorem, basic laws of statistics and the assumption of a memoryless channel according to (see, e. g., [17])

$$\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_n, \hat{\mathbf{x}}_0^{n-1}) = \frac{1}{C} \cdot \mathrm{P}(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)}) \cdot \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) \,, \qquad (2)$$

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

Prediction probabilities $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1})$ comprise *a priori* knowledge about a possibly transmitted bit combination $\mathbf{x}^{(i)}$ given the complete history of receive values $\hat{\mathbf{x}}_0^{n-1}$. For the computation of this term the following approaches are employed in this work.

2.1.1. Prediction Probabilities with Static 0th-Order A Priori Knowledge

If the transmitted bit combination can be modeled as an output of a 0th-order Markov process, then the prediction probability is given by $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) = \mathrm{P}(\mathbf{x}^{(i)}) = \mathrm{P}(s^{(i)})$, with $s^{(i)}$ being the sample value belonging to $\mathbf{x}^{(i)}$ [5, 18]. The probability $\mathrm{P}(s^{(i)})$ can be determined by a histogram measurement in a training process, leading to *a priori* knowledge of 0th order (AK0). However, larger Markov model orders exploiting the residual redundancy of multiple preceding sample values exhibit the problem of exponential rise of complexity and storage (see, e. g., [8, 19]). For example, a memory size of 2^{2M} is required for a 1st-order Markov model, which corresponds to approximately 8 GB for CD-quality audio with $M = 16$. Furthermore, approximately 2^{2M} multiply accumulates (MACs) have to be computed for every sample instant n (2^M MACs for the prediction probabilities for each of the 2^M possible values of $s^{(i)}$ during the final estimation). Naturally, this is technically not feasible for fine-quantized audio signals. As a result, we employ a newer sophisticated and efficient approach to prediction probabilities for high-quality audio, which is briefly described in the following.

2.1.2. Prediction Probabilities with Linear Predictive N_p-th-Order A Priori Knowledge

In order to efficiently exploit the residual redundancy in a large number of preceding estimated samples, we build upon a newer approach

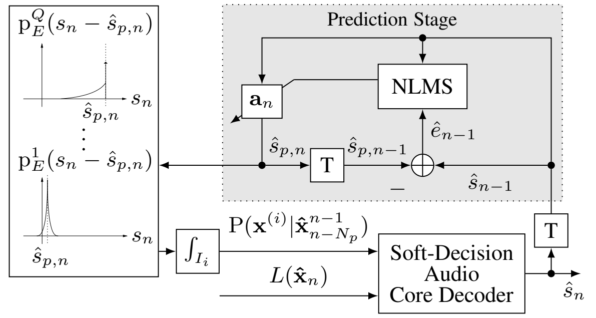


Fig. 3: Block diagram of the soft-decision audio decoder with prediction probabilities $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1})$ computed by linear prediction.

based on linear prediction presented in [8, 14, 19]. Here, we compute a prediction value $\hat{s}_{p,n}$ by a linear combination of N_p previously estimated sample values $\hat{\mathbf{s}}_{n-N_p}^{n-1} = (\hat{s}_{n-N_p}, \hat{s}_{n-N_p+1}, \ldots, \hat{s}_{n-1})^T$ according to

$$\hat{s}_{p,n} = \mathbf{a}_n^T \cdot \hat{\mathbf{s}}_{n-N_p}^{n-1} \,, \qquad (3)$$

with N_p denoting the prediction order, $\mathbf{a} = (a_n(N_p), a_n(N_p - 1), \ldots, a_n(1))^T$ being prediction coefficients and $(\cdot)^T$ denoting a transposed vector. Thereby, the influence of the previously estimated samples $\hat{\mathbf{s}}_{n-N_p}^{n-1}$ on the final estimate $\hat{s}_n$ can be condensed in $\hat{s}_{p,n}$ due to $\hat{s}_{p,n} = E\{\hat{s}_n|\hat{\mathbf{s}}_{n-N_p}^{n-1}\}$, with $E\{\cdot\}$ denoting the expected value. With a sufficiently large prediction order N_p and the direct association of $\mathbf{x}^{(i)}$ to $s^{(i)}$ and $\hat{\mathbf{x}}_{n-N_p}^{n-1}$ to $\hat{\mathbf{s}}_{n-N_p}^{n-1}$, respectively, the prediction probabilities can be approximated by

$$\begin{aligned} \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}) &\approx \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-N_p}^{n-1}) \\ &\approx \mathrm{P}(s^{(i)}|\hat{s}_{p,n}) \,. \end{aligned} \qquad (4)$$

Finally, the probability $\mathrm{P}(s^{(i)}|\hat{s}_{p,n})$ for a possibly transmitted sample value $s^{(i)}$ given the prediction value $\hat{s}_{p,n}$ can be determined by shifting the probability density function (PDF) $\mathrm{p}_E(\hat{e}_n = \hat{s}_n - \hat{s}_{p,n})$ of the prediction error $\hat{e}_n$ by $\hat{s}_{p,n}$ (see Fig. 3) and integration over the i-th linear PCM quantization interval[1] I_i according to

$$\mathrm{P}(s^{(i)}|\hat{s}_{p,n}) = \int_{I_i} \mathrm{p}_E(s_n - \hat{s}_{p,n}) \, \mathrm{d}s_n \,. \qquad (5)$$

We employ the well-known normalized least-mean-square (NLMS) algorithm (see, e. g., [20]) for a dynamic update of prediction coefficients after each estimation. The NLMS algorithm is also widely utilized in the field of lossless audio encoding [21, 22]. The corresponding NLMS adaptation rule can be written as

$$\mathbf{a}_{n+1} = \mathbf{a}_n + \frac{\hat{e}_n}{1 + \lambda \cdot ||\hat{\mathbf{s}}_{n-N_p}^{n-1}||^2} \cdot \hat{\mathbf{s}}_{n-N_p}^{n-1} \,, \qquad (6)$$

with λ being a tuning parameter controlling the convergence rate [21]. At sample index $n = 0$, the variables are initialized according to $\mathbf{a}_{-1} = (1/N_p, \ldots, 1/N_p)^T$ and $\hat{\mathbf{s}}_{-N_p}^{-1} = (0, \ldots, 0)^T$.

The decoding results can be improved if information about the current predictability of the audio signal is exploited. Therefore, multiple adaptively-shaped prediction error PDFs $\mathrm{p}_E^q(e)$

[1] The integration intervals are given by $I_0 = (-\infty, -1]$, $I_i = [-1 + (i-1) \cdot \Delta, -1 + i \cdot \Delta]$, for $i \in \{1, 2, \ldots, 2^M - 2\}$, and $I_{2^M-1} = [1 - 2 \cdot \Delta, \infty)$.

are employed, $q \in \{1, 2, \ldots, Q\}$, each being trained for different magnitude intervals of the prediction value, i. e., $q = \min(\lfloor Q \cdot |s_{p,n}| + 1 \rfloor, Q)$. A comprehensive description of this approach has been presented in [19].

2.2. Audio Sample Estimation

The received sample is finally estimated with the *a posteriori* probabilities from (2) by employing the minimum mean-square error (MMSE) $E\{(s_n - \hat{s}_n)^2\} \to \min.$ as an error criterion. The corresponding estimation rule can be written as

$$\hat{s}_n = \sum_{i=0}^{2^M-1} s^{(i)} \cdot \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_{n-Np}^{n}) \,. \tag{7}$$

3. UTILIZATION OF EXPLICIT REDUNDANCY IN SOFT-DECISION AUDIO DECODING

The Bayesian framework in (2) allows us to exploit any available *a priori* knowledge about a possibly transmitted bit combination $\mathbf{x}^{(i)}$ for the estimation of audio samples. As a result, it enables us to utilize explicit redundancy that has been appended to the information bits $\mathbf{x}_n$ on the transmitter side, e. g., by means of sample-individual block codes. Consequently, a transmitter-sided channel encoder is necessary, yielding a code word $\mathbf{y}_n$ for every $\mathbf{x}_n$ that is to be transmitted over the equivalent channel. In order to maintain a clear presentation, we limit the following description to systematic code words $\mathbf{y}_n = (\mathbf{x}_n, \mathbf{z}_n)$, with $\mathbf{z}_n \in \{\mathbf{z}^{(0)}, \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(r)}, \ldots, \mathbf{z}^{(2^R-1)}\}$ being a transmitted parity bit combination, $\mathbf{z}^{(r)}$ being a possible parity bit combination of a rate ${}^{M}\!/_{M+R}$ block code and R denoting the total number of parity bits (cf. [5, 14]).

As a result, analog to the derivation of (2) the required *a posteriori* probability $\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^n, \hat{\mathbf{z}}_0^n)$ can be split up by applying Bayes' theorem, basic laws of statistics and marginalization over r according to

$$\begin{aligned}\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^n, \hat{\mathbf{z}}_0^n) = \frac{1}{C} \cdot \mathrm{P}(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)}) \cdot \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}, \hat{\mathbf{z}}_0^{n-1}) \\ \cdot \sum_{r=0}^{2^R-1} \mathrm{P}(\hat{\mathbf{z}}_n|\mathbf{z}^{(r)}) \cdot \mathrm{P}(\mathbf{z}^{(r)}|\mathbf{x}^{(i)}) \,,\end{aligned} \tag{8}$$

with $\hat{\mathbf{z}}_n$ being the received hard-decided parity bits, $\mathrm{P}(\hat{\mathbf{z}}_n|\mathbf{z}^{(r)})$ denoting the transition probability of the parity bits, and $\mathrm{P}(\mathbf{z}^{(r)}|\mathbf{x}^{(i)})$ being a (pseudo-)statistical description of the employed block code. The term $\mathrm{P}(\mathbf{z}^{(r)}|\mathbf{x}^{(i)})$ must be known in advance and can be stored in the receiver as a $2^R \times 2^M$-dimensional matrix.

Let us discuss an example: For a single-parity-check code $(\mathbf{x}^{(i)}, z^{(r)})$ with $R = 1$ Eq. (8) reads as

$$\begin{aligned}\mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^n, \hat{\mathbf{z}}_0^n) =& \frac{1}{C} \cdot \mathrm{P}(\hat{\mathbf{x}}_n|\mathbf{x}^{(i)}) \cdot \mathrm{P}(\mathbf{x}^{(i)}|\hat{\mathbf{x}}_0^{n-1}, \hat{\mathbf{z}}_0^{n-1}) \\ &\cdot \Big(\mathrm{P}(\hat{z}_n|z^{(0)} = 0) \cdot \mathrm{P}(z^{(0)} = 0|\mathbf{x}^{(i)}) \\ &+ \mathrm{P}(\hat{z}_n|z^{(1)} = 1) \cdot \mathrm{P}(z^{(1)} = 1|\mathbf{x}^{(i)}) \Big) \,.\end{aligned} \tag{9}$$

In the case that this parity check covers only the two most significant bits (MSBs, gray-shaded) of a natural-binary-mapped bit combination with $M = 3$ bits, the corresponding parity-check matrix can be given in pseudo-statistical form by

$$\mathrm{P}(z^{(r)}|\mathbf{x}^{(i)}) = \begin{array}{l} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{array}{l} \scriptstyle r=0:\ z_n=0 \\ \scriptstyle r=1:\ z_n=1 \end{array} \end{array}$$

$$\text{with} \quad \begin{array}{rcccccccccl} & & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & \text{LSB} \\ \mathbf{x}^{(i)} = & & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & \\ & & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \text{MSB} \\ i = & & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \end{array} \,.$$

The implementation of more advanced sample-individual block codes with $R > 1$ is straightforward by utilizing (8).

4. SIMULATIONS

4.1. Experimental Setup

We evaluate the proposed explicit redundancy approach by carrying out simulations with 13 monaural audio signals transmitted over an equivalent channel disturbed by additive white Gaussian noise (AWGN) with binary phase-shift keying (BPSK) and coherent soft-demodulation (cf., e. g., [23]). All audio signals are pulse-code modulated (PCM), sampled at 48 kHz, initially uniformly quantized with 24 bits per sample (as reference signals for quality measures), then with 16 bits per sample, and are normalized to −26 dBFS (decibels relative to full-scale). The total length of the audio signals is 96 s and they comprise excerpts from classical pieces and a motion-picture soundtrack with effects and instruments like organs, brass instruments, strings, percussions, pianos and synthesizers. For an increased reliability of our measurements all files are transmitted five times for every investigated channel state.

The *a priori* probabilities $\mathrm{P}(\mathbf{x}^{(i)})$ and prior distributions $\mathrm{p}_E^q(e)$ required for the approaches presented in Secs. 2.1.1 and 2.1.2, respectively, have been measured in a training process from a database of 15 musical pieces with a total length of 81 min (comprising various pieces of classical and electronic music, and a motion-picture soundtrack with speech, music and effects) exclusively utilized for training purposes.

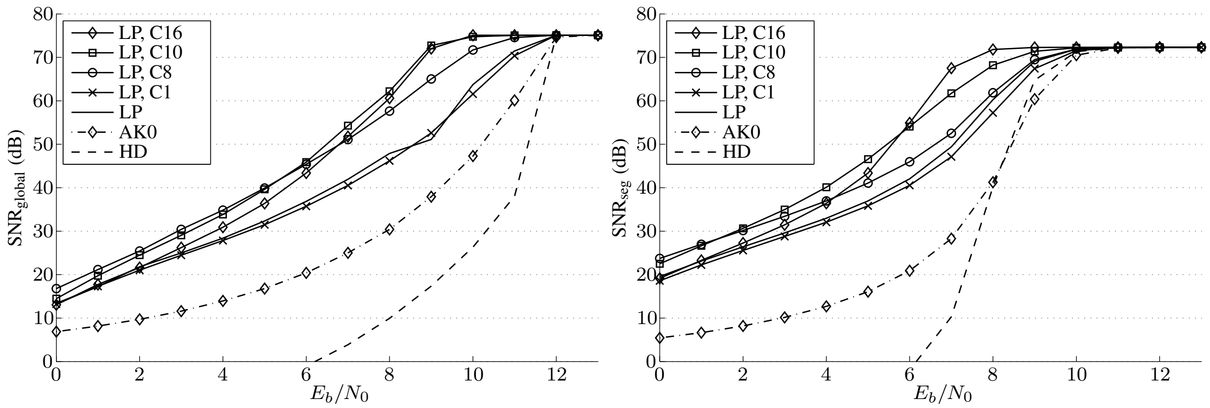
In order to measure the decoding performance, we employ global ($\mathrm{SNR}_{\mathrm{global}}$) and segmental ($\mathrm{SNR}_{\mathrm{seg}}$) signal-to-noise ratio (SNR) measurements (cf., e. g., [24]), and the basic model implementation of the perceptual evaluation of audio quality (PEAQ) standard [25, 26] as described in [27]. For every measurement, the reference signal $\tilde{s}$ quantized with 24 bits per sample and the estimated 16 bit audio sample $\hat{s}$ are available. The segmental SNR is computed with a segment size of 480 samples (i. e., 10 ms). The SNR measures are especially useful in the field of error concealment, due to $\mathrm{SNR}_{\mathrm{global}}$ being sensitive to rare transient distortions like high-energy clicks, whereas $\mathrm{SNR}_{\mathrm{seg}}$ being sensitive to permanent but low-amplitude artifacts like crackling. The PEAQ Recommendation defines *objective difference grades* (ODGs) ranging from 0.0 (imperceptible) to −4.0 (very annoying)[2].

Experiments carried out in the preparatory stage showed that the parameters $\lambda = 20$ for the NLMS algorithm, a prediction order of $N_p = 10$ and a number of $Q = 16$ adaptively-shaped prior distributions are reasonable choices.

4.2. Discussion

In order to evaluate the effects of employing explicit redundancy for delayless soft-decision audio decoding, simulations with single-parity-check ($R = 1$) codes C1, C2, …, C16 covering 1, 2, …,

[2]Please note that the algorithm given in the PEAQ Recommendation actually leads to a maximum ODG score of 0.22.



(a) Global SNR (b) Segmental SNR

Fig. 4: SNR performance of soft-decision decoding based on linear prediction with $N_p = 10$, $Q = 16$, and parity codes Cx covering a number of x MSBs (LP, Cx), soft-decision decoding based on linear prediction with $N_p = 10$, $Q = 16$, and without explicit redundancy (LP), with 0th order *a priori* knowledge (AK0) and hard-decision decoding (HD).

	$\mathbf{E_b/N_0}$					
Method	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB
LP, C16	-3.8	-3.5	-2.5	-1.3	$\mathbf{0.1}$	$\mathbf{0.2}$
LP, C10	-3.8	-3.3	-2.2	$\mathbf{-1.0}$	$\mathbf{0.1}$	$\mathbf{0.2}$
LP, C8	$\mathbf{-3.7}$	$\mathbf{-3.2}$	$\mathbf{-2.1}$	-1.3	-0.2	$\mathbf{0.2}$
LP, C1	-3.8	-3.6	-3.0	-2.1	-1.0	0.1
LP	-3.8	-3.6	-2.9	-1.9	-0.8	0.1
AK0	-3.9	-3.9	-3.8	-3.9	-3.5	-1.2
HD	-3.9	-3.9	-3.9	-3.9	-3.2	-2.7

Table 1: Mean PEAQ ODG scores of soft-decision decoding based on linear prediction with $N_p = 10$, $Q = 16$, and parity codes Cx covering a number of x MSBs (LP, Cx), soft-decision decoding based on linear prediction with $N_p = 10$, $Q = 16$, and without explicit redundancy (LP), with 0th order *a priori* knowledge (AK0) and hard-decision decoding (HD). Best results are printed in boldface.

16 MSBs of the transmitted bit combinations $\mathbf{x}_n$, respectively, have been carried out. In order to maintain an accessible presentation, the most important results are presented in Fig. 4 and Table 1 for the linear prediction approach with explicit redundancy (LP, Cx), without explicit redundancy (LP), soft-decision decoding with 0th order *a priori* knowledge (AK0) and hard-decision decoding (HD). Please note that the total energy of the transmitted *coded* bit combinations $(\mathbf{x}_n, z_n)$ is equal to the total energy of the *uncoded* bit combinations $\mathbf{x}_n$, enabling a direct and fair comparison of all depicted schemes.

It can be seen that the exploitation of explicit redundancy in terms of sample-individual parity-check codes for PCM audio signals leads to enormous gains in $\mathrm{SNR_{global}}$, $\mathrm{SNR_{seg}}$ and PEAQ ODG, if the parity bit considers a sufficiently large number of MSBs. Regarding the maximum achievable audio SNR of approximately 75 dB, the investigated codes notably improve the decoding results by 2 dB E_b/N_0 compared to the corresponding LP approach without explicit redundancy.

The C16 code covering whole bit combinations $\mathbf{x}_n$ significantly improves the decoding results by approximately 20 dB ($\mathrm{SNR_{global}}$ and $\mathrm{SNR_{seg}}$) compared to the LP approach for medium channel quality in the range of $E_b/N_0 \approx 7 \ldots 8$ dB. However, the SNR and PEAQ ODG performance worsens enormously for lower E_b/N_0 ratios. This is due to the low significance of the parity bit when too many disturbed bits are considered in the decoding process.

In contrast, the employment of the C1 code covering only one MSB has negative effects on the decoding results, worsening the $\mathrm{SNR_{global}}$ results by a few decibels compared to the uncoded LP approach. This shows that a sufficient amount of information bits needs to be included into the parity checks in order to yield noticeable gains in audio quality.

The application of the C8 and C10 codes leads to the most balanced decoding results. The C8 code delivers the best SNR values for $E_b/N_0 = 0$ dB and the C10 code for E_b/N_0 ratios between 2 dB and 5 dB. Furthermore, the C10 code greatly improves the $\mathrm{SNR_{seg}}$ performance for $E_b/N_0 > 5$ dB, without the performance loss of the C16 code for extremely bad channel states. This is also reflected by the PEAQ ODG results.

5. CONCLUSIONS

In this contribution we have evaluated the employment of explicit redundancy for robust soft-decision decoding of high-quality audio signals. The explicit redundancy in terms of sample-individual block codes is being exploited delaylessly within the underlying Bayesian framework utilized for soft decoding. Besides *explicit* redundancy, channel reliability information in terms of LLRs and *implicit* redundancy within the audio signal are exploited. Simulation results with single-parity-check codes on an AWGN channel with BPSK modulation show a considerable gain in audio quality compared to soft decoding relying only on inherent redundancy and reliability information. Precisely, for the case of a nearly artifact-free audio transmission, the utilization of explicit redundancy by a single parity bit leads to a gain of 2 dB in E_b/N_0. Regarding audio SNR and PEAQ ODG, gains of approximately 20 dB and 0.9 ODG scale units, respectively, can be observed. Our approach can be applied to transmission systems for high-quality audio with very high demands on latency and robustness, e. g., wireless digital microphones.

6. REFERENCES

[1] C. E. Shannon, "Communication in the Presence of Noise," *Proc. I.R.E.*, vol. 37, pp. 10–21, Jan. 1949.

[2] 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.

[3] 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.

[4] J. Hagenauer, "Source-Controlled Channel Decoding," *IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2449–2457, Sep. 1995.

[5] 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.

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

[7] W. Xu, J. Hagenauer, and J. Hollmann, "Joint Source-Channel Decoding Using the Residual Redundancy in Compressed Images," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, Dallas, TX, U.S.A., Jun. 1996, pp. 142–148.

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

[9] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, U.S.A.: Prentice Hall, 1982.

[10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[11] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, vol. 3, Dallas, TX, U.S.A., Nov. 1989, pp. 1680–1686.

[12] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, U.S.A.: The MIT Press, 1963.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding And Decoding: Turbo-Codes," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[14] F. Pflug and T. Fingscheidt, "Robust Ultra-Low Latency Soft-Decision Decoding of High-Quality Audio," *IEEE Trans. Audio, Speech, Language Process.*, pp. 1–12, Sep. 2012, submitted for publication.

[15] T. Fingscheidt, S. Heinen, and P. Vary, "Joint Speech Codec Parameter and Channel Decoding of Parameter Individual Block Codes (PIBC)," in *Proc. IEEE Workshop Speech Coding*, 1999, pp. 75–77.

[16] M. C. Valenti and S. Cheng, "Iterative Demodulation and Decoding of Turbo-Coded M-ary Noncoherent Orthogonal Modulation," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1739–1747, Sep. 2005.

[17] T. Fingscheidt, "Parameter Models and Estimators in Soft Decision Source Decoding," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds. Chichester, West Sussex, England: Wiley, 2008, pp. 281–310.

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

[19] F. Pflug and T. Fingscheidt, "Delayless Soft-Decision Decoding of High-Quality Audio with Adaptively Shaped Priors," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Barcelona, Spain, Sep. 2011, pp. 2099–2103.

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

[21] 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.

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

[23] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA, U.S.A.: McGraw-Hill, 2001.

[24] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, A. V. Oppenheim, Ed. Eaglewood Cliffs, NJ, U.S.A.: Prentice-Hall, 1984.

[25] "Method for Objective Measurements of Perceived Quadio Quality," Rec. ITU-R BS.1387-1, Nov. 2001.

[26] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *J. Audio Eng. Soc.*, vol. 48, no. 1, pp. 3–29, Feb. 2000.

[27] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality," Dept. Electrical & Comp. Eng., McGill University, Montreal, QC, Canada, TSP Lab Technical Report, May 2002.