AN SBR TOOL FOR VERY LOW DELAY APPLICATIONS WITH FLEXIBLE CROSSOVER FREQUENCY

Michael Werner †

Gerald Schuller †*

† Ilmenau University of Technology
Helmholtzplatz 2, 98693 Ilmenau, Germany
werner.michael@tu-ilmenau.de

* Fraunhofer IDMT
Ehrenbergstrasse 29
98693 Ilmenau, Germany

ABSTRACT

In the field of audio coding a technique called Spectral Band Replication (SBR) is used in several codecs to reduce the data rate. We already developed an SBR tool with very little algorithmic delay for use in low delay applications like teleconferencing or live music performance but it produces a relatively high amount of side information. Further it only supports an SBR crossover frequency at half the signal bandwidth. This paper presents an enhanced SBR tool for low delay applications which uses a method known from speech coding called Codebook Mapping (CBM) to get a reduced side information data rate and an optional 3 band decomposition to reach low crossover frequencies. This leads to a low delay SBR tool with a reduced amount of side information and flexible crossover frequencies.

***Index Terms*—** low delay audio coding, bandwidth extension, spectral band replication, codebook mapping

1. INTRODUCTION AND PREVIOUS APPROACHES

In the last years the demand for audio codecs for very time-critical applications like wireless microphones, in ear monitoring or video conferencing has increased. But apart from a low delay and a good sound quality, in many cases these codecs have to operate at small data rates.

One popular approach to decrease the bit rate of coded audio is called Spectral Band Replication (SBR) [1, 2]. The main principle of this technique is the assumption that there exists a correlation between the lower and the upper audio spectrum. To save bits, the audio signal is split into a lower and higher band, and only the lower band is coded with a common audio coder. The higher band is described by a small number of parameters which are transmitted as side information. In the decoder the higher band is reconstructed out of a copy of the decoded lower band using the side information. This approach is successfully applied in the MPEG HE-AAC codec [3]. Because the MPEG-SBR module uses a 64 channel QMF decomposition to obtain the higher band parameters, its algorithmic delay reaches about 30 ms for a 32 kHz signal (regardless of the HE-AAC core delay). This is too much for many time-critical applications which require a delay less than 10 ms.

The proposed low delay SBR is based on a previous version of a low delay SBR tool (prev-2BS, [4]). In that system the higher band is reconstructed using LPC parameters describing the original higher band envelope. Because these parameters have to be transmitted for every signal block and because of the small block size used in that system the amount of side information reaches about 4,71 kbps. An additional disadvantage of that system is, that it only works with a crossover frequency at half of the signal bandwidth.

In speech coding there exists an alternative technique to regenerate the higher band of a band limited audio signal, which is called Codebook Mapping (CBM) [5, 6, 7]. The main aspect of this method is a one-to-one relation between the lower band spectrum and the higher band spectrum of a speech signal. By connecting typical pairs of spectral lower-/ higher band envelopes in a codebook, the higher band can be reconstructed in the decoder by only searching the lower band codebook. No transmission of side information is necessary.

Our goal for this paper is to apply the CBM technique to obtain a system which reduces the data rate of low delay audio codecs while keeping a high sound quality and a very low delay.

2. NEW APPROACH

Our new approach is to combine the techniques used in prev-2BS and extend them by additional crossover frequencies and CBM.

The central idea is to use CBM to find the optimal higher band envelope parameters in the decoder instead of transmitting it as side information. In contrast to the simple CBM principle we utilize the advanced CBM method shown in figure 1: While in a simple CBM system every lower band codebook entry is associated exactly to one higher band envelope, in our system every lower band codebook entry has several possible higher band envelopes assigned to it. These are called embedded codebooks. Because the embedded codebooks are highly adapted to their particular lower band entry, the reconstructed higher band will reach a better quality than the simple CBM [8]. The drawback of this system is the additional side information caused by the codebook indices (4 bit in our system) which are needed in the decoder to find the correct higher band entry.

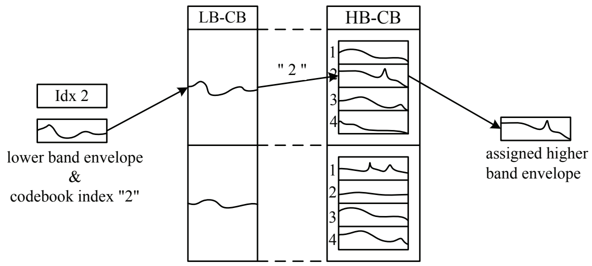

Fig. 1. Codebook Mapping with embedded codebooks: Every single entry in the lower band codebook (LB-CB) has an individual higher band codebook (HB-CB) assigned to it.

Additionally the new system can operate with a crossover frequency not only at half the original signal bandwidth, but also at a third. When the input audio signal is resampled to a suitable sampling rate, this results in an increased flexibility to obtain a desired crossover frequency and bandwidth for the reconstructed higher band. Because of the multiple codebook search procedures our new system has a slightly increased complexity compared to prev-2BS.

2.1. Two Band System (2BS)

ENCODER. The 2BS shown in figure 2 works with two frequency bands which have the same bandwidth. Hence the standard sampling rates 48 kHz, 44.1 kHz, 32 kHz and 22.05 kHz correspond to the crossover frequencies 12 kHz, 11.025 kHz, 8 kHz and 5.5 kHz. To get a certain crossover frequency the input signal has to be resampled to the corresponding sampling rate first. After downsampling each subband by a factor of two they are blocked in 50% overlap-add frames $X_{HB/LB}(m)$ with a length of 128 samples. Afterwards the prediction coefficients of each higher band and lower band frame and the prediction error power $E_{HB}(m)$ of the higher band are estimated using a LPC analysis. These coefficient vectors $A_{HB/LB}(m)$, which describe the lower band envelope, are compared with the different coefficient vectors in a precalculated lower band codebook. The squared error distortion measure [9], because of its low complexity, is utilized for finding the best matching codebook entry. Based on the estimated higher band coefficients $A_{HB}(m)$ the corresponding higher band envelope is then determined in the embedded higher band codebook, which has 16 entries. The 4 bit index of this higher band codebook entry is transmitted in addition to the 6 bit logarithmic quantized prediction error power $E_{HB}(m)$ of the higher band. The prediction error power is needed in the decoder to scale the reconstructed higher band to its correct power (hence the term scale factor). In total the side information of the 2BS adds up to 10 bit per frame.

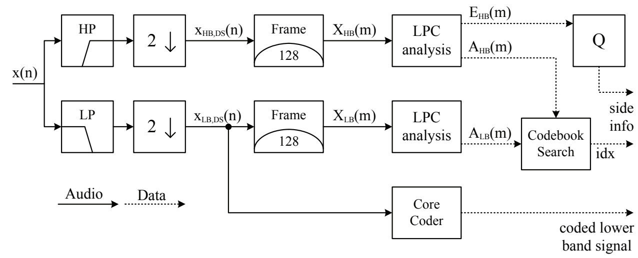

Fig. 2. The encoder of the Two Band System (2BS).

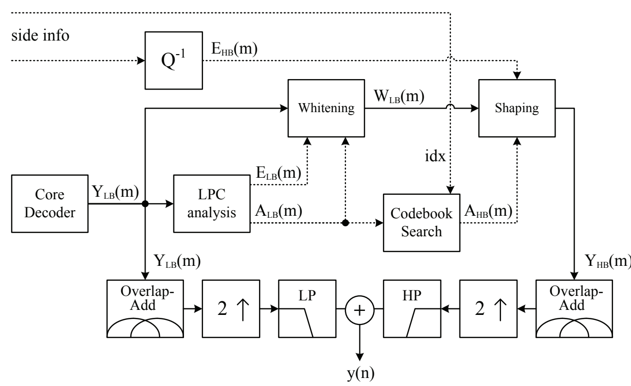

Fig. 3. The decoder of the Two Band System (2BS).

DECODER. In the decoder depicted in figure 3, first the spectral envelope of the decoded lower band is determined by a LPC analysis. The low band signal has the same 128-samples frame structure as in the encoder. The result is one coefficient vector $A_{LB}(m)$ per frame that describes the lower band envelope. In the same way as in the encoder the best matching lower band codebook entry is identified by the codebook search. With help of the transmitted codebook index idx it is now possible to find the best higher band envelope $A_{HB}(m)$. Simultaneously the lower band signal $Y_{LB}(m)$ passes the whitening filter. Here it is spectrally flattened by using a predictive analysis filter what can be seen as filtering the lower band signal with its inverse envelope and gain. For this operation the lower band envelope $A_{LB}(m)$ and its scale factor $E_{LB}(m)$ are used. The output is the signal $W_{LB}(m)$ with variance 1. In the shaping filter this signal gets the envelope and the gain of the original higher band. For this a predictive synthesis filter uses the just found higher band prediction coefficients $A_{HB}(m)$ to shape the spectral envelope of the signal $W_{LB}(m)$. Then it is scaled to the variance of the original higher band utilizing the transmitted scale factor $E_{HB}(m)$. At the end the complete signal $y(n)$ is created by upsampling, filtering and adding the lower band and higher band.

2.2. Three Band System (3BS)

The 3BS has the same structure as the 2BS with the only difference that it works with three frequency bands instead of two. All three bands have the same bandwidth, which results in possible crossover frequencies of 8 kHz, 7.3 kHz, 5.3 kHz and 3.7 kHz when using the standard sampling rates. The three band CBM works as follows: Every single entry in the lower band codebook has an order-16 middle band and higher band codebook assigned to it. Thus, in the decoder it is possible to recreate the middle band and the higher band with only one transmitted codebook index and two scale factors (one for the middle band, one for the higher band). All the other operations are basically identical to the 2BS.

3. CODEBOOK DESIGN

The codebook pairs were calculated based on a 30 minute test sequence of music, ambient noise and speech signals using the popular LBG algorithm [9]. We created codebooks with 1024 lower band entries and 16 entries in the ebmedded codebook. This number of codebook entries is a tradeoff between sound quality, search complexity and bit rate. Further details can be found in [8].

4. SOUND QUALITY

The 2BS and the 3BS were tested against our prev-2BS and against the standard MPEG-SBR in MUSHRA listening tests [10]. To compare the performance under extreme conditions the crossover frequency of all SBR systems was set to about 5.5 kHz. To get that crossover frequency in our system, 12 MPEG test files were resampled to 32 kHz and 22.05 kHz. After that they were coded with the different systems listed in table 1 and presented to the probands who had to rate the particular versions. In the tests, no core coder was used because we only wanted to evaluate the SBR tools themselves. The results are shown in figure 4. One can see that the overall average scores for our systems are slightly below MPEG-SBR. Our prev-2BS and the 2BS and 3BS are all located in the "good" range. The reduced quality of speech is probably caused by overemphasized peaks in the reconstructed middle / higher band which could be suppressed in future systems. Observe that for the signals "Pop Music", "Harpsichord" and "Plucked Strings" the 3BS is evaluated significantly better than the 2BS. This is probably caused by more high frequency components in those signals which are better reconstructed by the 3BS because of the higher sampling frequency of that system.

	Stimulus	Parameter
1	**Reference**	unprocessed, $f_s = 32kHz$
2	**Anchor 3.5 kHz**	lowpass filter at 3.5 kHz
3	**Anchor 7 kHz**	lowpass filter at 7 kHz
4	**2BS**	$f_s = 22.05kHz$, $f_c = 5513Hz$
5	**3BS**	$f_s = 32kHz$, $f_c = 5333Hz$
6	**prev-2BS**	$f_s = 22.05kHz$, $f_c = 5513Hz$
7	**MPEG-SBR**	$f_s = 32kHz$, $f_c = 5500Hz$

Table 1. Stimuli of the MUSHRA listening tests with sampling frequency f_s and crossover frequency f_c.

5. DATA RATE

In the 2BS the side information consists of one 6 bit scale factor and one 4 bit higher band codebook index per frame (16 higher band entries). That results in 1.68 kbps for a sampling rate of 22.05 kHz. The 3BS needs two scale factors, one per frame for the middle band and one for the higher band, and one 4 bit codebook index. The result is a side information data rate of 3.91 kbps for a sampling frequency of 32 kHz.

6. DELAY

The modules which cause a delay are the filter bank and the overlap-add blocking scheme. For the two channel filter bank we used Cauer filters of order 16, which produce a delay of 5 samples. The three channel filter bank utilizes cauer filters as well, but to reach a comparable narrow transition band steepness the bandpass had to be designed with order 36, which creates a delay of 16 samples. In the 2BS, the 128 samples blocking scheme of the downsampled signal causes a delay of 256 samples of the broadband signal. Including the delay of the filter banks the overall delay sums up to 266 samples. Because of the downsampling by a factor of three in the 3BS, the blocking scheme causes a delay of 384 samples in the broadband signal. With the filter banks delay the total delay is 416 samples. In table 2 all details of the tested SBR systems are listed.

	sampl. freq. [kHz]	cross. freq. [Hz]	data rate [kbps]	delay [samples]
2BS	22.05	5513	1.68	266
3BS	32	5333	3.91	416
prev-2BS	22.05	5513	2.36	266
MPEG-SBR	32	5500	3.20	961

Table 2. Sampling frequency, crossover frequency, side information data rate and algorithmic delay of the different SBR tools as configured in the listening test.

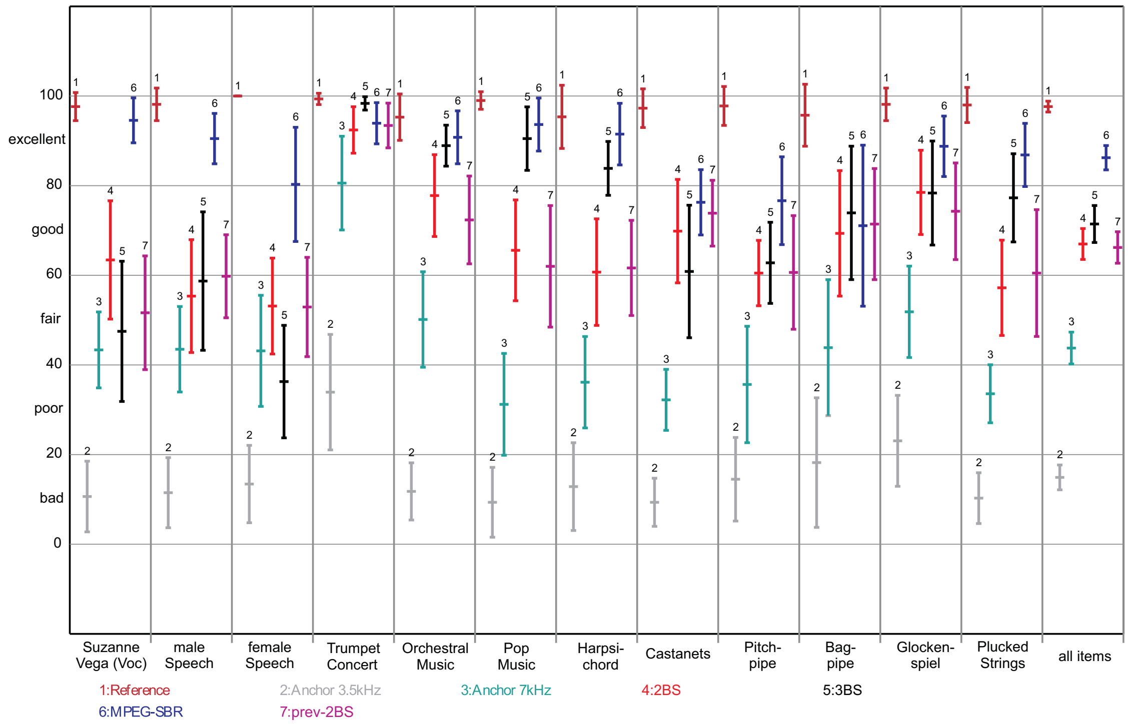

Fig. 4. Mean Subjective Score with 95% confidence intervals for the MUSHRA listening test session (14 listeners).

7. CONCLUSION

Both proposed variants of our low delay SBR tool are an improvement compared to prev-2BS: The 2BS delivers a comparable sound quality and a lower side information data rate while keeping the delay low. Compared to the 2BS the 3BS can reproduce a higher signal bandwidth but at the expense of a slightly increased delay and a somewhat higher amount of side information. The listening test showed significant improvements for signals with strong high frequency components. In comparison to MPEG-SBR our systems show a slightly reduced sound quality with the advantage of less than only one third of the delay and a similar side information data rate.

8. REFERENCES

[1] L. G. Liljeryd et al., “Source Coding Enhancement Using Spectral-Band Replication,” United States Patent No. US 7283955 B2, October 2003

[2] M. Dietz et al., “Spectral Band Replication, a novel approach in audio coding,” AES 112th convention, Munich, Germany, May 2002

[3] 3GPP, “Enhanced aacPlus encoder SBR part (TS 26.404 V7.0.0 Release 7),” 3GPP technical specification, June 2007

[4] T. Friedrich and G. Schuller, “Spectral Band Replication Tool for Very Low Delay Audio Coding Applications,” Proc. IEEE WASPAA, October 2007, 199-202

[5] J. Epps and W. Holmes, “A new technique for wideband enhancement of coded narrowband speech,” Proc. IEEE Workshop on Speech Coding, 1999, 174-176

[6] J. A. Fuemmeler, R. C. Hardie, W. R. Gardner, “Techniques for the Regeneration of Wideband Speech from Narrowband Speech,” EURASIP Journal on Applied Signal Processing, 2001, 266-274

[7] E. Larsen, R. M. Aarts. *Audio bandwidth extension: application of psychoacoustics, signal processing and loudspeaker design*. Wiley, 2004

[8] M. Werner, G. Schuller, “An enhanced SBR Tool for Low Delay Applications,” AES 127th convention, New York, USA, October 2009

[9] A. Buzo et al., “Speech coding based upon vector quantization,” IEEE Transactions on Acoustics, Speech and Signal Processing 28, October 1980, Nr. 5, 562-574

[10] ITU-R, “Recommendation BS.1534-1: Method for the subjective assessment of intermediate audio quality,” International Telecommunication Union, Geneva, Switzerland, 2001