LOW-COMPLEXITY SEMI-PARAMETRIC JOINT-STEREO AUDIO TRANSFORM CODING

Christian R. Helmrich[1], *Andreas Niedermeier*[2], *Stefan Bayer*[1], *and Bernd Edler*[1]

[1] International Audio Laboratories Erlangen [2] Fraunhofer Institut für Integrierte Schaltungen (IIS)
Am Wolfsmantel 33, 91058 Erlangen, Germany
christian.helmrich@audiolabs-erlangen.de

ABSTRACT

Traditional audio codecs based on real-valued transforms utilize separate and largely independent algorithmic schemes for parametric coding of noise-like or high-frequency spectral components as well as channel pairs. It is shown that in the frequency-domain part of coders such as Extended HE-AAC, these schemes can be unified into a single algorithmic block located at the core of the modified discrete cosine transform path, enabling greater flexibility like semi-parametric coding and large savings in codec delay and complexity. This paper focuses on the stereo coding aspect of this block and demonstrates that, by using specially chosen spectral configurations when deriving the parametric side-information in the encoder, perceptual artifacts can be reduced and the spatial processing in the decoder can remain real-valued. Listening tests confirm the benefit of our proposal at intermediate bit-rates.

Index Terms — Audio coding, decorrelation, MDCT, stereo

1. INTRODUCTION

Following a need for unifying the previously separate speech and transform coding paradigms into a single general-purpose audio codec (coder/decoder), three new codec standards were developed during the last decade. Building upon the principle introduced in Extended AMR-WB [1], two new codecs were proposed in 2012: Extended HE-AAC based on AMR-WB+ and HE-AAC v2 [2], and Opus based on SILK and CELT [3]. The respective transform-coding parts of these two standards, namely the improved HE-AAC v2 and CELT, both employ the modified discrete cosine transform (MDCT) [4] to obtain frequency-domain (FD) representations for quantization and coding of a frame. They also provide three parametric tools:

• *noise filling* with pseudo-random MDCT coefficients. This is called anti-collapse (and to some extent, spreading) in CELT and perceptual noise substitution (PNS) in legacy HE-AAC.

• *high-frequency reconstruction* for bandwidth extension. This is known as spectral band replication (SBR) in HE-AAC. In CELT a similar effect can be achieved with spectral folding.

• *parametric stereo* for efficient joint coding of two channels. This is represented by traditional intensity stereo [5] in CELT and by more elaborate MPEG Surround 2-1-2 coding in [2].

Whereas CELT's folding and stereo tools operate directly in the MDCT domain, SBR and MPEG Surround require additional pseudo-QMF banks for analysis and synthesis, which increase not only the audio quality but also algorithmic delay and complexity (both by roughly a factor of two). Given the limited battery power of mobile equipment, it is desirable to minimize the complexity of codecs running on such devices, in our case by avoiding the pseudo-QMF banks while, hopefully, retaining the high audio quality achieved with them. We also aim for semi-parametric coding, i. e. mixing of parametric and MDCT-based coding in a spectral band. Due to separate domains, this is impossible with SBR or MPEG Surround.

Several methods for high-frequency reconstruction (HFR) or parametric stereo (PS) with high quality, operating directly on the MDCT coefficients of an audio signal, have been published over the last years, most recently by Lee and Choi [6], Neukam et al. [7], Sheng et al. [8], Suresh and Raj [9], Tammi et al. [10], Tsujino et al. [11], and Zhang et al. [12]. Most of these methods apply various more or less fundamental structural changes to the underlying MDCT-based coders in order to improve the audio quality. Moreover, only [10] addresses semi-parametric coding, but does so in a scalability context.

To this end we developed a "minimally invasive" unified tool for HFR and PS which can easily be integrated into any MDCT-based audio codec without necessitating architectural modifications. This tool, by not requiring any auxiliary filter banks other than the existing MDCT, allows semi-parametric coding with very low complexity and no additional delay. An earlier, less unified version of the tool is adopted in [13]. The general concept of the HFR component, called intelligent gap filling (IGF), is described in other publications [14, 15]. Here we focus on the stereo algorithms of the IGF and PS parts.

The remainder of this paper is organized as follows. Section 2 illuminates where semi-parametric coding is beneficial, and Section 3 describes how noise filling, IGF, and PS can be combined and integrated into a single functional block in an Extended HE-AAC decoder to enable such semi-parametric coding. The detailed operation of the stereo methods on both decoder and encoder side is discussed in Section 4, outlining a scheme devised to minimize temporal artifacts caused by the underlying real-valued transform. Section 5 then presents and examines the results of tests conducted to evaluate the subjective and objective performance of our proposal in comparison to the state of the art. Finally, Section 6 concludes the paper.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

2. PARAMETRIC CODING IN HE-AAC AND CELT

HE-AAC and Extended HE-AAC, like Opus, allow efficient coding of an audio waveform by segmenting it into frames via overlapping windows, transforming each window to a spectral domain using a MDCT, and jointly quantizing and coding the resulting MDCT bins in frequency (and sometimes also temporal) groups. These frequency groups, which are called scale factor bands in HE-AAC and energy bands in CELT, closely follow the psychoacoustic Bark or ERB scales [16] in terms of their bandwidths and represent the domain in which the following basic tools are operated in both encoder and decoder:

- discrete, that is invertible, mid-side (sum-difference) stereo, enhanced by complex prediction in Extended HE-AAC [17],
- in HE-AAC and CELT, intensity stereo using downmixing,
- noise filling or anti-collapse of MDCT bins quantized to 0.

Implementing these tools, of which the latter two are parametric and the former two can be enabled selectively for each band, is straightforward and well-known. More sophisticated HFR and PS coding, however, are not trivial especially when high audio quality is requested. In particular, aliasing artifacts are likely to occur when operating in the real-valued MDCT based on time-domain aliasing cancelation (TDAC) [18]. For this reason, both HFR coding (via SBR) and PS coding (via MPEG Surround) are located in the complex-valued domain of a pseudo-QMF bank in (Extended) HE-AAC. A general block diagram of the (Extended) HE-AAC coding-decoding chain with its QMF and MDCT tools is shown in Figure 1.

Fig. 1 indicates that in the encoder, the MDCT core-coder has to wait for the result of the QMF-based pre-processing, while in the decoder, the QMF tools require the output of the MDCT core-decoder. Moreover, the core signal is downsampled. Since the "inner" MDCT and "outer" SBR and MPEG Surround codecs operate in separate filter-bank domains, such sequential processing leads to accumulation of the individual domain's algorithmic delays. In fact, the sum of all delays is more than 200 ms at 44.1 or 48 kHz input sample rate, while the core-codec alone exhibits a delay of roughly 120 ms [19]. Semi-parametric coding is prevented as well, as shown below.

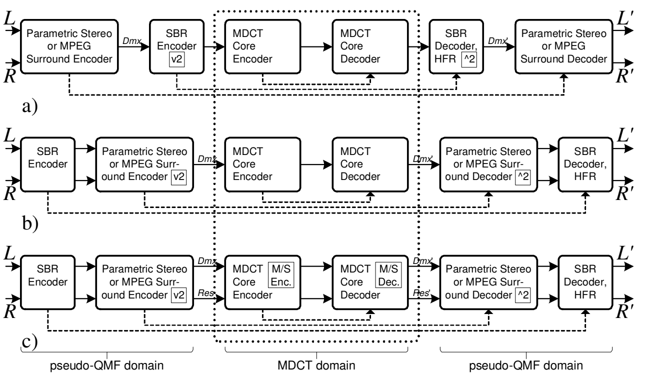

Fig. 1. *Block diagram showing (–) signal and (--) side-information flow in (Extended) HE-AAC. The dotted box encapsulates the core.*

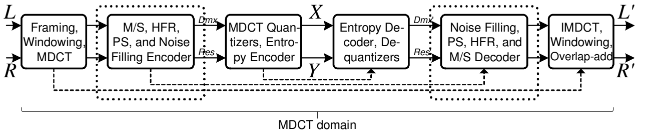

Fig. 2. *Block diagram of the proposed modification to the HE-AAC system of Fig. 1. The dotted boxes mark the location of the SBS tool.*

2.1. Principle and Benefit of Semi-Parametric Coding

At some bit-rates, neither fully parametric nor fully waveform preserving "discrete" coding deliver satisfactory quality when employing (Extended) HE-AAC – SBR and MPEG Surround saturate below a sufficiently high level of quality[1], while pure MDCT-based coding requires too coarse (i. e. limited-quality) spectral quantization to match the bit-consumption target. At such intermediate bit-rates, it was found necessary to support both parametric and discrete coding within the same spectral region, adapted on a frame-by-frame basis, to increase fidelity. Such mixing of paradigms, termed semi-parametric coding in this paper, is reflected by the MDCT-based noise filling tools in Extended HE-AAC and CELT, where within a scale factor (or energy) band B at index k, quantized spectral coefficients X_i and inserted noise values X_j may coexist without overlap,

$$B(k) = X_i + X_j, \quad i \in \omega_k, \quad j \in \nu_k, \tag{1}$$

that is, with $\omega_k \cap \nu_k = \varnothing$. In SBR and MPEG Surround, such an approach is very difficult to achieve without deteriorating the sonic quality of MDCT-coded components overlapping with the parametric-coding part, since the pseudo-QMF sub-bands exhibit lower frequency resolution than the MDCT bins. For this reason, MDCT content in MPEG Surround or SBR coded regions is set to zero in a (Extended) HE-AAC decoder. Since a removal of the SBR and PS coding concepts is problematic from a quality perspective, it is desirable to move their parametric coding functionalities into the MDCT domain, as noted earlier, in order to enable simple semi-parametric coding.

It is worth noting that SBR contains its own noise addition tool to counteract excessive tonality or spectral holes in the HFR region resulting from translating (copying up) core-coder spectral regions which are tonal or largely quantized to zero [20]. Thus, two noise insertion tools exist in HE-AAC, one in SBR and the other in the MDCT codec. In the following section we propose a modification to the scheme depicted in Figure 1 which removes the redundancy in the noise filling tools and at the same time unifies all three parametric methods discussed so far. Additional filter-bank domains other than the one to which the MDCT belongs are therefore not required.

3. UNIFIED MDCT-BASED PARAMETRIC CODING

The only solution to the issues outlined in the last section is to replace the SBR and MPEG Surround tools by alternative realizations operating directly in the MDCT domain of the FD core-coder. A respective proposal is shown in Figure 2. Essen-

1 due to band-limiting of the waveform preserving core being downsampled by a factor of two

tially, this approach reduces the "unified stereo" configuration of Fig. 1 c (MPEG Surround downmixing with optional band-limited residual) to a core-codec-only version, with the former two *outer* parametric tools merged into the *inner* M/S stereo and noise filling blocks. As a result, the total system delay is reduced to less than 60 ms, given that additional processing look-ahead or downsampling are not used any more. This unified MDCT-based stereo and parametric coding tool, in which a single noise filling tool is shared (thus avoiding the noted redundancy), is named *spectral band substitution* (SBS).

A notable advantage of the modified scheme of Fig. 2 over those of Fig. 1 is increased spectral flexibility: the encoder is given the possibility to apply either parametric (i.e. waveform destructive) or discrete (i. e. waveform preserving) coding per scale factor band. Due to the sequential ordering in different domains, as seen in Fig. 1, this is not supported in (Extended) HE-AAC. A consequence is that, by employing the scheme of Fig. 2, three particular decisions can be taken by the encoder:

- no static, pre-determined cross-over frequencies at which HFR coding starts or at which residual coding in PS ends; this could be set individually per frame based on the input.
- bandwise choice of parametric or discrete coding based on the load of the entropy coder trying to reach the bit-rate; if it runs out of bits, parametric coding is used in more bands.
- band-wise use of parametric tools based on psychoacoustic criteria; a resp. model could control the activation of HFR and PS based on how well these tools represent each band.

The third advantage is of particular interest since it can be used to avoid both tonality mismatch which can occur in SBR as well as destructive interference which can happen in MPEG Surround in the downmix process. This will be examined further in the next section. It must be noted that SBR and MPEG Surround can also be configured so that the core-encoder does not run out of bits. This, however, can only be done a priori.

A close-up diagram of the proposed SBS decoding algorithm is depicted in Figure 3. The SBS system could, of course, also be integrated into legacy AAC, CELT, or any other transform codec using spectral bands, but for the sake of brevity, seamless integration, and efficient re-use of existing tools like complex prediction [17], the remainder of this paper focuses on an integration into Extended HE-AAC. Moreover, only the noise filling and stereo aspects of SBS will be described, since HFR by way of IGF has already been documented in [14, 15].

4. STEREO CODING DESIGN AND OPTIMIZATION

The unified SBS system consists of 3 consecutive tools – noise filling using a pseudo-random source, PS by means of stereophonic frequency filling (Steffi), and IGF – all of which are applied on the joint-stereo coded MDCT channel bins X, Y on a per-frame basis, as illustrated in Fig. 3. The noise filling tool is identical to the one in [2], particularly in its use only above 0.08 times the sample rate f_S (3.45 kHz at f_S = 44.1 kHz), so it won't be detailed here. For clarification it is emphasized that only those samples of X (or Y) quantized to zero are filled with noise N, scaled by a noise level $0 \leq l_N < 1$, as reflected by (1):

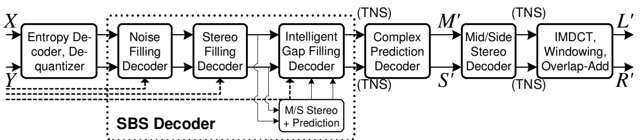

Fig. 3. *Detailed procedure of spectral band substitution decoding in the context of an implementation into the Extended HE-AAC decoder.*

$$X_j = l_N \cdot N_j, \;\; j \in v_k, \;\; N_j \in \{-1, 1\}. \tag{2}$$

The Steffi tool is closely linked with the noise filling tool and, thus, inherits the conservative start frequency of $0.08 \cdot f_S$ to avoid low-frequency artifacts and the restriction to operate only on the set v_k of MDCT samples within band k. It further allows for relatively fine mixing with the noise filling output by way of l_N, as clarified in the next subsection. Once X and Y have been noise and/or stereo filled, IGF can be employed for HFR by filling high-frequency samples quantized to zero (and thus constituting spectral gaps) using non-zero low-frequency content. The output of the SBS tool-chain is finally processed by the other FD tools available in Extended HE-AAC, namely Temporal Noise Shaping (TNS) and M/S stereo with complex prediction [2, 17], whose order in the decoder can be signaled.

4.1. Detailed Decoder Design and Integration

For maximum flexibility and minimum complexity overhead, SBS is tightly integrated into the Extended HE-AAC decoder. After noise filling of both channels, Steffi is used in the side (i. e. second) spectrum Y to "fill up" all scale factor bands k at or above $0.08 \cdot f_S$ which were fully quantized to zero, using the corresponding spectral bins of the downmix obtained from the preceding frame's decoded L' and R' spectra (see also Fig. 3):

$$S_j = Y_j + l_M(k) \cdot D(M'_j), \;\; j \in v_k \wedge \omega_k = \varnothing, \tag{3}$$

with M' being the fully decoded downmix and $D(\cdot)$ specifying the frame-delay operation. This simple approach is both convenient and efficient because M' is already computed for the complex stereo prediction tool in Extended HE-AAC [2, 17]. Moreover, combined with M/S decoding, (2) and (3) create a Lauridsen-type decorrelator [21] since an upmix of M' and a frame-delayed version of itself is equivalent to a pair of complementary FIR comb filters. This can easily be demonstrated if a stationary two-channel white-noise spectrum $\{L, R\}$ to be reconstructed via Steffi is assumed. In that case the downmix is $M = (L + R) / 2$, the residual $S = (L - R) / 2$ is not transmitted but reproduced at the decoder side as $S' = D(M')$, and through M/S decoding, $\{L', R'\}$ is derived via addition and subtraction,

$$L' = M' + S' = M' + D(M'), \tag{4}$$
$$R' = M' - S' = M' - D(M'), \tag{5}$$

assuming $l_N = 0$ and $l_M = 1$. This is a simplified version of the decorrelators in HE-AAC v2 [22] or MPEG Surround [2]. The transfer function of the processes (2)–(5) is shown in Figure 4.

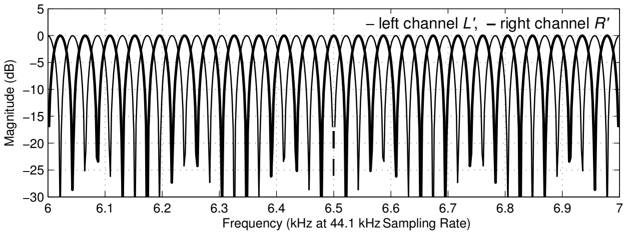

Fig. 4. *Transfer function of the decorrelator resulting from the simple example of jointly applying Steffi and M/S decoding with l_N=0, l_M=1.*

Like IGF, Steffi produces a target energy $E_T(k)$ in band k by adding to k decoded MDCT bins from another spectral region. However, instead of spectral *copy-up* from lower frequencies, temporal *copy-over* from the last frame's MDCT downmix is applied. More precisely, a core (coded) energy $C(k)$ is derived:

$$C(k) \;=\; \sum_j Y_j^2, \;\; j \in \nu_k, \tag{6}$$

with Y_j constituted as specified in (2). Given the source energy

$$E_M(k) \;=\; \sum_j D(M'_j)^2, \;\; j \in \nu_k, \tag{7}$$

with $D(M'_j)$ as in (3), and the fact that Y_j and $D(M'_j)$ generally are uncorrelated – since the former is either zero or randomly noise-filled – $l_M(k)$ in (3) can simply be calculated as follows:

$$l_M(k) \;=\; \sqrt{\frac{w_k \;-\; C(k)}{E_M(k) \;+\; \varepsilon}}, \tag{8}$$

where w_k denotes the number of MDCT coefficients in k [14] and ε is a tiny constant to avoid divisions by zero. To prevent energy collapse when $D(M'_j)$ is zero, the energy of S_j is found after execution of (3) and corrected in case it differs from w_k:

$$E_S(k) \;=\; \sum_j S_j^2, \;\; S'_j \;=\; S_j \cdot \sqrt{\frac{w_k \cdot E_k^2}{E_S(k) \;+\; \varepsilon}}, \;\; E_k^{\;2} \;=\; \frac{E_T(k)}{w_k}. \tag{9}$$

S'_j, subjected to the stereo prediction and M/S decoders (4, 5), exhibits energy $E_T(k)$ due to proper scaling by E_k, which can be coded as a legacy AAC scale factor for k since $\omega_k = \varnothing$.

Note that the copy-up process in IGF may be conducted on partially jointly coded X and Y in case of stereophonic coding, so some source bins for IGF band b might be left-right coded while b shall always be mid-side coded (or downmix-residual coded in case of stereo prediction). In such cases, it is necessary to first convert the effective L' and R' source bins for the copy-up into a M' and S' representation, as depicted in Fig. 3.

4.2. Encoder Design and Optimization

A block diagram of the SBS encoder transmitting the spectra X, Y and Steffi side-information l_N, E_k is presented in Figure 5. Details on each component are omitted for brevity. Three aspects, however, shall be emphasized. First, bins not quantized to zero in the SBS spectral region are not affected by SBS. We exploit this to retain, i.e. quantize to non-zero, bins of specific spectral components in the IGF or Steffi regions which cannot be represented well by the copy-fill approach. Particularly, the core-coder acts similarly to SBR's "missing harmonics" tool [20] in the SBS bands, albeit in a waveform preserving way. Second, by quantizing a Steffi band in Y to non-zero when its energy greatly exceeds that of the respective downmix band in X, destructive interferences in the output signal – and notch or comb filter artifacts associated with them – are reduced. Third, instead of simply measuring the MDCT energy in each Steffi band of S before its quantization to obtain E_k, complex-valued energy ratios involving MDST values, as in equations (13–16) in [14], can be used to minimize temporal modulation artifacts (E_k for IGF in [14] has the same meaning as E_k in this paper).

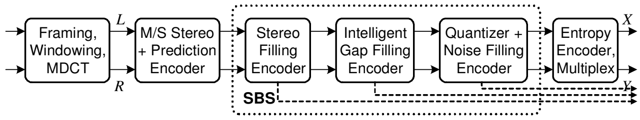

Fig. 5. *Detailed SBS encoder scheme, spectral composition proposal.*

5. IMPLEMENTATION AND EVALUATION OF SBS

Two subjective tests and an objective analysis were conducted in order to assess the performance of the SBS proposal when integrated into Fraunhofer's Extended HE-AAC encoder and decoder. The subjective evaluation comprised listening tests in accordance with the MUSHRA (*mul*tiple *s*timuli with *h*idden *r*eference and *a*nchor) method [23], performed by 12 listeners (max. age 37) via Stax headphones, the objective analysis was a complexity estimation. In all assessments, the codec was run at f_S = 48 kHz in a "low complexity" configuration (to disable SBR and MPEG Surround), 16.5 kHz bandwidth, and real (not complex) stereo prediction in the SBS spectral region, as this proved to be the best tradeoff between quality and complexity.

5.1. Listening Test without MDCT Quantization

The first blind test intended to evaluate the perceptual merit of Steffi over simple noise filling of Y or "unified stereo" MPEG Surround 2-1-2 (20 parameter bands) and was executed without MDCT quantization or IGF. The test material consisted of 7 critical stereo items previously utilized during MPEG audio standardizations, plus 10 s of pink noise to check for particular decorrelation artifacts. Figure 6 a) presents the results.

Three aspects can serve to explain why Steffi offers a level of quality which always matches, and sometimes exceeds, that of the unified stereo scheme. First, discrete waveform coding was employed below 750 Hz in both codecs under test, so any advantages of MPEG Surround due to its more sophisticated decorrelation were not revealed. Second, given that a unified stereo frame has 2048 time-samples, while Steffi operates on the 1024-sample core-frames, the mean side-information rate of MPEG Surround (2 kbit/s) – and accordingly the time resolution – was lower than that of Steffi (3–4 kbit/s). Third, the encoder optimizations noted in subsection 4.2 seem to work.

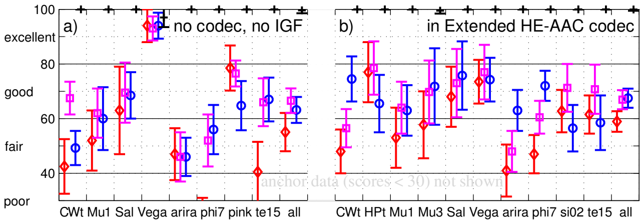

Fig. 6. *Results of the two MUSHRA tests conducted to evaluate SBS.* (♦) *only noise filling,* (■) *SBS,* (●) *unified stereo ref.,* (+) *hidden ref.*

5.2. Listening Test with MDCT Quantization

The second listening test examined the merit of the Steffi and IGF combination in a typical application scenario of 48 kbit/s Extended HE-AAC coding, i. e. with enabled FD quantization and noise filling. The results illustrated in Figure 6 b) reveal the small but significant overall improvement due to the SBS component but also slight quality shortcomings for three items compared to the MPEG Surround and SBR coded reference. This is caused by a lack of phase coding in Steffi and the quite high side-information rate of SBS, leaving less bit-budget for MDCT core-coding. On average, however, the performance of the reference codec is matched by the SBS proposal, which is satisfactory given the complexity difference, as shown below.

5.3. Evaluation of Computational Complexity

A decoder's algorithmic complexity, as noted in the introduction, is a critical aspect especially on battery-powered mobile devices, so an estimation of the workload of an Extended HE-AAC decoder with SBS or MPEG Surround + SBR was made with an ARM 926 simulator. By virtue of its efficient integration, SBS caused the peak stereo-core decoder load to increase by only 9 % to about 22 MHz. This is only one half of a mono-core HE-AAC v2 decoder's and one third of a unified stereo Extended HE-AAC decoder's workload, both of which utilize 30 extra MHz per core channel for QMF-domain processing.

6. CONCLUSION

In this paper we proposed a spectral band substitution (SBS) technique, working directly in the MDCT domain of an audio codec such as HE-AAC and unifying (and by doing so, representing an alternative to) the previously separated noise filling and the QMF-domain SBR and parametric stereo (PS) tools. Particular parameter calculation in the encoder was shown to ensure consistent quality of the decoded signal without having to resort to complex-valued processing in the decoder. Formal subjective evaluation of SBS in the Extended HE-AAC codec indicates that the high level of audio quality offered by QMF-based PS methods is reached at a fraction of their algorithmic delay and complexity, at least for bit-rates at which semi-parametric coding is useful. Studies at lower bit-rates will follow.

7. ACKNOWLEDGMENT

The authors thank Michael Fischer for help in evaluating SBS.

REFERENCES

[1] R. Lefebvre et al., "Extended AMR-WB for High-Quality Audio on Mobile Devices," *IEEE Comm. Mag.*, vol. 44, no. 5, 2006.
[2] ISO/IEC 23003-3:2012, "MPEG Audio Technologies – Part 3: Unified Speech and Audio Coding," Geneva, Jan. 2012.
[3] J. M. Valin, K. Vos, and T. Terriberry, "Definition of the Opus audio codec," IETF RFC 6716. http://tools.ietf.org/html/rfc6716
[4] J. P. Princen, A. W. Johnson, and A. B. Bradley, "Subband/transform coding using filter bank design based on time domain aliasing cancellation, *Proc. IEEE ICASSP-87*, vol. 12, April 1987.
[5] R. G. van der Waal and R. N. J. Veldhuis, "Subband coding of stereophonic digital audio signals," in *Proc. IEEE ICASSP-91*, Toronto, vol. 5, pp. 3601–3604, April 1991.
[6] Y. H. Lee and S. H. Choi, "Superwideband Bandwidth Extension using Normalized MDCT Coefficients for Scalable Speech and Audio Coding," *Advances in Multimedia*, no. 909124, 2013.
[7] C. Neukam, F. Nagel, G. Schuller, and M. Schnabel, "A MDCT based harmonic spectral bandwidth extension method," in *Proc. IEEE ICASSP-13*, Vancouver, pp. 566–570, May 2013.
[8] C. Sheng, H. Ruimin, and P. Yuxing, "Spatial parameter choosing method based on spatial perception entropy judgment," in *Proc. 8th Int. Conf. on WiCOM*, Shanghai, pp. 1–4, Sep. 2012.
[9] K. Suresh and A. R. Raj, "MDCT domain parametric stereo audio coding," *Proc. Int. Conf. on SPCOM*, Bangalore, pp. 1–4, 2012.
[10] M. Tammi, L. Laaksonen, A. Rämö, and H. Toukomaa, "Scalable superwideband extension for wideband coding," in *Proc. IEEE ICASSP-09*, Taipei, pp. 161–164, April 2009.
[11] K. Tsujino and K. Kikuiri, "Low-complexity bandwidth extension in MDCT domain for low-bitrate speech coding," in *Proc. IEEE ICASSP-09*, Taipei, pp. 4145–4148, April 2009.
[12] S. Zhang, W. Dou, and H. Yang, "MDCT Sinusoidal Analysis for Audio Signals Analysis and Processing," *IEEE. Trans. on Audio, Speech, and Language Proc.*, vol. 21, no. 7, July 2013.
[13] ISO/IEC 23008-3, "MPEG-H – Part 3: 3D Audio," Mar. 2015.
[14] C. R. Helmrich, A. Niedermeier, S. Disch and F. Ghido, "Spectral envelope reconstruction via IGF for audio transform coding," in *Proc. IEEE ICASSP-15*, Brisbane, April 2015.
[15] S. Disch, C. Neukam, and K. Schmidt, "Temporal tile shaping for spectral gap filling in audio transform coding in EVS," in *Proc. IEEE ICASSP-15*, Brisbane, April 2015.
[16] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 6th edition, Bingley, UK: Emerald Group Publishing, 2012.
[17] C. R. Helmrich et al., "Efficient transform coding of two-channel audio signals by means of complex-valued stereo prediction," in *Proc. IEEE ICASSP-11*, Prague, pp. 497–500, May 2011.
[18] F. Küch and B. Edler, "Aliasing reduction for modified discrete cosine transform domain filtering and its application to speech enhancement," in *Proc. IEEE WASPAA*, pp. 131–134, Oct. 2007.
[19] M. Lutzky et al., "A guideline to audio codec delay," in *Proc. AES 116th Convention*, Berlin, no. 6062, May 2004.
[20] A. C. den Brinker et al., "An Overview of the Coding Standard MPEG-4 Audio Amendments 1 and 2: HE-AAC, SSC, and HE-AAC v2," *EURASIP Journal on ASMP*, no. 468971, June 2009.
[21] C. Gribben and H. Lee, "The perceptual effects of horizontal and vertical interchannel decorrelation using the Lauridsen decorrelator," *Proc. AES 136th Convention*, Berlin, no. 9027, April 2014.
[22] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric Coding of Stereo Audio," *EURASIP Journal on Applied Signal Processing*, vol. 9, no. 561917, June 2005.
[23] ITU, Radiocommunication Assembly, "Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems," Geneva, Jan. 2003.