SIGNAL-ADAPTIVE SWITCHING OF OVERLAP RATIO IN AUDIO TRANSFORM CODING

Christian R. Helmrich and Bernd Edler

International Audio Laboratories Erlangen (Friedrich-Alexander University of Erlangen-Nürnberg)
Am Wolfsmantel 33, 91058 Erlangen, Germany
christian.helmrich@audiolabs-erlangen.de

ABSTRACT

Contemporary perceptual audio coders, all of which apply the modified discrete cosine transform (MDCT), with an overlap ratio of 50%, for frequency-domain quantization, provide good coding quality even at low bit-rates. However, relatively long frames are required for acceptable low-rate performance also for quasi-stationary harmonic input, leading to increased algorithmic latency and reduced temporal coding resolution. This paper investigates the alternative approach of employing the extended lapped transform (ELT), with 75% overlap ratio, on such input. To maintain a high time resolution for coding of transient segments, the ELT definition is modified such that frame-wise switching between ELT (for quasi-stationary) and MDCT coding (for non-stationary or non-tonal regions), with complete time-domain aliasing cancelation and no increase in frame length, becomes possible. A new ELT window function with improved side-lobe rejection to avoid framing artifacts is also derived. Blind subjective evaluation of the switched-ratio proposal confirms the benefit of the signal-adaptive design.

***Index Terms*—** audio coding, lapped transform, MDCT

1. INTRODUCTION

During the last 22 years, particularly since the development of the MPEG-1 Layer 3 (MP3) and AC-3 (Dolby Digital) coders, perceptual audio coding has relied exclusively on the modified discrete cosine transform (MDCT), introduced by Princen *et al.* [1, 2] and further investigated, under the name modulated lapped transform (MLT), by Malvar [3], for waveform preserving spectral quantization. The inverse of this transform, given a length-M spectrum X'_i for frame index i, can be written as

$$x'_i(n) = \frac{2}{M} \sum_{k=0}^{M-1} X'_i(k) \cos\Big(\tfrac{\pi}{M}\big(n + \tfrac{M+1}{2}\big)\big(k + \tfrac{1}{2}\big)\Big), \quad (1)$$

with $0 \le n < N$ and N being the window length. Since $M = \frac{N}{2}$, the overlapping ratio is 50%. In recent standards based on the MPEG-2 Advanced Audio Coding (AAC) specification [4, 5], this concept has been extended to also allow parametric tools such as noise filling in the MDCT domain. The new MPEG-H 3D Audio framework [6, 7], for example, offers the following functionality for semi-parametric transform-domain coding:

- noise filling of zeroed spectral lines above some frequency,
- stereo filling for semi-parametric joint-stereo coding [8, 9],
- Intelligent Gap Filling (IGF) for bandwidth extension [10].

In [9], the combination of IGF and stereo filling, entitled spectral band substitution (SBS) in [8], assisted by transform kernel switching for input with non-trivial inter-channel phase differences, was shown to deliver *good* audio quality for most signals. On quasi-stationary harmonic segments, however, the subjective performance was lower than that of the alternative high-delay/complexity 3D Audio configuration using spectral band replication (SBR) and "unified stereo" MPEG Surround in a pseudo-QMF domain. An explanation for this behavior is the higher frequency resolution of the MDCTs utilized in the latter configuration: at the given output sample rate of 48 kHz, the M-size core transforms operate on 24-kHz downsampled downmix and residual signals, doubling the frame length.

SBS-based 3D Audio coding, due to its delay, complexity, and temporal-resolution advantages [8], represents the variant of choice at least for mono- and stereophonic signals, and it is desirable to improve its design – while maintaining the frame length – such that its performance can match that of the QMF-based configuration even on single-instrument and other tonal recordings. A viable solution for increased spectral efficiency on quasi-stationary segments is the extended lapped transform (ELT) proposed by Malvar [11, 12], whose inverse (synthesis) version is identical to (1), except that $0 \le n < L$ with $L \ge 4M$. Unfortunately, as will be shown in section 2, its overlap ratio is at least 75% instead of the MDCT's 50%, which often leads to audible artifacts for transient waveform parts like drum hits or tone onsets. Moreover, practical solutions for *block length switching* between ELTs of different lengths – or between an ELT and MLT – similarly to the technique applied in MDCT codecs for precisely such transient frames, have not been presented (only theoretical work has been published [13–17]).

To address this shortcoming, section 3 proposes a simple modification to the ELT formulation of (1), allowing perfectly reconstructing transitions between transforms with 50% and 75% overlap ratio, along with a novel ELT window. Section 4 then introduces a signal-adaptive coding scheme applying the switched-ratio principle. The designs and results of subjective tests of this scheme, integrated into the 3D Audio codec, are discussed in section 5. Finally, section 6 concludes the paper.

2. PRINCIPLES OF LAPPED TRANSFORM CODING

The ELT, MLT, and MDCT, as indicated in section 1, can be considered specific realizations of a general lapped transform formulation, with (1) for the inverse and with $0 \le k < M$ and

$$X_i(k) = \sum_{n=0}^{L-1} x_i(n)\ \mathrm{cs}\Big(\tfrac{\pi}{M}\big(n + \tfrac{M+1}{2}\big)\big(k + \tfrac{1}{2}\big)\Big) \qquad (2)$$

for the forward (analysis) case. Notice that the cos( ) function has been replaced by placeholder cs( ) in (2) to stress that one may also use the sin( ) function in (1, 2) to obtain sine modulated forms like the modified discrete sine transform (MDST) applied in the MCLT [18] and the authors' prior work [8, 9].

To attain perfect reconstruction (PR) of input signal $s_i(n)$ after subjection to analysis and synthesis transforms (1, 2), at least in the absence of spectral distortion e.g. by quantization (indicated by a $'$ in (1)), windows $w(n)$ are used to weight the L-size analysis input $x_i(n) = w(n) \cdot s_i(n)$ as well as the synthesis output $\hat{s}_i(n) = w(n) \cdot x_i'(n)$. Since $\hat{s}_i(n)$ exhibits time domain aliasing (TDA) due to the critical-sampling property of lapped transformation, $w(n)$ must fulfill particular design constraints [1, 2, 12]. For ELTs with even $\frac{L}{M}$, assuming equal, symmetric $w(n)$ for analysis and synthesis, these are given by

$$\sum_{j=0}^{\frac{L}{M}-2l-1} w(k + jM)w(k + jM + 2lM) = \delta(l),\ \ 0 \le l < \tfrac{L}{2M}. \qquad (3)$$

For the MLT, MDCT, or MDST ($\frac{L}{M} = \frac{N}{M} = 2$, the three terms will apply interchangeably hereafter), the TDA is canceled by combining the first temporal half of $\hat{s}_i$ with the second half of the previous frame's $\hat{s}_{i-1}$ by means of an overlap-and-add (OLA) procedure. The resulting inter-transform overlap ratio is $\frac{2-1}{2}$. In case of the ELT with $L = 4M$, the OLA step must combine the first quarter of $\hat{s}_i$ with the second quarter of $\hat{s}_{i-1}$, the third quarter of $\hat{s}_{i-2}$, and the fourth quarter of $\hat{s}_{i-3}$, so the ratio grows to $\frac{4-1}{4}$. Figure 1 illustrates this difference and the worst-case pre-echo (temporal spread of coding errors). More detailed discussions of TDA and PR can be found in [15–20]. Note, also, that evenly stacked linear-phase ELTs based on the DCT-II, or odd-length ELTs with e.g. $L = 3M$, are possible as well [21, 22], but such designs will not be examined here.

Focusing on the length-$4M$ ELT ($\frac{L}{M} = 4$), one can observe that, as shown in Figure 2a, PR is not achieved during switchovers to and from MLT coding since the TDA symmetries are incompatible – simply put, the necessity of adjacent even-odd combinations [9, 19] is violated between frames $i-4$ and $i-3$.

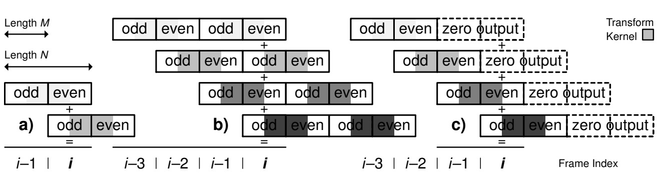

Fig. 1. *TDA cancelation during OLA in lapped transformation for* (a) *MLT,* (b) *ELT,* (c) *MLT via ELT.* (-) *Maximum pre-echo.*

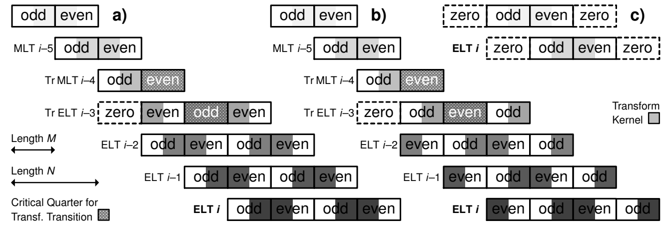

Fig. 2. *Switch from MLT to ELT with* (Tr)*ansition transforms:* (a) *incorrect non-PR,* (b) *desired PR.* (c) *MLT via desired ELT.*

3. A MODIFIED EXTENDED LAPPED TRANSFORM

To correct the PR issue in Fig. 2a by achieving complete TDA cancelation (TDAC) also in the transitory 3-part OLA regions, one transform class must be redefined such that its TDA symmetries complement those of the other, e.g. as in Figs. 2b, 2c. Since it is desirable to avoid changes to existing MDCT and MDST implementations, focus is laid on the ELT. Moreover, to easily obtain PR transition and steady-state windows for all transforms, corresponding analytic expressions are necessary.

3.1. Modifications for Adaptation of Overlap Ratio

To give the ELT the desired TDA compatibility with the MLT, it suffices to alter the temporal phase shift in its base functions:

$$\hat{X}_i(k) = \sum_{n=0}^{L-1} x_i(n)\ \mathrm{cs}\Big(\tfrac{\pi}{M}\big(n + \tfrac{3M+1}{2}\big)\big(k + \tfrac{1}{2}\big)\Big), \qquad (4)$$

with k, cs as in (2) and the inverse ELT (1), using $\hat{X}_i'$, adapted accordingly. It can also be shown that, as Fig. 2 indicates, the 4 quarters of the transitory MLT and ELT windows are based on the respective steady-state weightings, with the first and/or fourth quarter set to zero and the critical quarters described by

$$w_{\mathrm{tr}}(t) = \sqrt{1 - w_{\mathrm{elt}}(k)^2 - w_{\mathrm{elt}}(M+k)^2}, \qquad (5)$$

where $t = \frac{L}{2} + k$ for switching as in Fig. 2 or $t = \frac{L}{2} - 1 - k$ for the reverse ELT-to-MLT transitions. Using (5) to acquire the critical quarters (white text in Fig. 2) for both the ELT and MLT transition weightings completes the definition of the transitory windows, leaving just the choice of the steady-state functions.

3.2. Steady-State PR Lapped-Transform Windows

Several power-complementary (PC) MLT windows enforcing the so-called Princen-Bradley condition for PR [2] have been documented. Figure 3a depicts the shapes and corresponding oversampled transfer functions of the windows used in MPEG audio codecs [5, 7], the MLT sine [3, 11] and the Kaiser-Bessel derived (KBD) windows [23]. Also shown is the PC function developed by the first author in [24], whose shape is similar to that of the KBD window but which, as can be noted, exhibits lower first (near-field) side-lobe levels. Finally, a sine window for a doubled frame length, as employed in case of dual-rate SBR, serves as a reference and illustrates that longer windows can notably reduce both pass-band width and stop-band level.

Ideally, an ELT window, subject to the PR constraints of (3), should exhibit a frequency response comparable to that of the double-length sine window, but it can be observed that, due to the PR restrictions, main-lobe width can only be minimized by allowing less side-lobe attenuation. Malvar's window [11] with $p{=}1$, for example, was found to have the lowest possible main-lobe width of all ELT designs but also undesirably high stop-band levels, as shown in Figure 3b. Its temporal borders are notably discontinuous (since samples beyond the window extension are assumed to equal zero), resulting in a side-lobe decay of only –6 dB/octave [24] and framing artifacts in our experiments. Temerinac and Edler [16] presented a recursive design approach, which they used to obtain the ELT window also shown in Fig. 3 (note that the value –0.038411 is missing in column "L=4N" of their table 1). This window, which can be closely approximated by Malvar's equations with $p{=}0.14$, provides more but still quite weak stop-band attenuation.

It is worth noting that, for $p{=}1$, Malvar's formulation can be simplified to a notation similar to that for a Hann window:

$$w_{p=1}(t) = a_0 - 0.5\cos\Big(2\pi \cdot \tfrac{t+0.5}{L}\Big),\ \ L = 4M, \tag{6}$$

with $0 \leq t < L$ denoting the temporal samples of the window and $a_0 = 2^{-3/2}$ chosen to enforce the PR constraints [11–14]. Intuitively, a function with more side-lobe attenuation such as

$$w_{\text{3-term}}(t) = \sum_{k=0}^{2} b_k \cos\Big(2k\pi \cdot \tfrac{t+0.5}{L}\Big),\ \ b_1 = -0.5, \tag{7}$$

with $b_2 > 0$, which can be used to derive Blackman's window [24], would seem applicable as well. Unfortunately, it can be shown that PR cannot be achieved with such a window class regardless of the value of b_0. However, by adding more terms,

$$w_{\text{elt}}(t) = w_{\text{3-term}}(t) - \sum_{k=1}^{K} c_k \cos\Big(8k\pi \cdot \tfrac{t+0.5}{L}\Big), \tag{8}$$

with b_k as above, the resulting shape for any choice of $b_2 \lesssim \frac{3}{8}$ can be corrected so that PR is approached arbitrarily closely. Targeting, in particular, a low stop-band level and imposing, in addition to the PR conditions, the restriction of an isotone left-half and, hence, antitone right-half window slope, PR can be approximated with an error below $4 \cdot 10^{-6}$ by using $K = 3$, $b_2{=}0.176759$ and, dependent on these values, $b_0{=}0.3303$ and

$$c_1 = 0.02366318,\ \ c_2 = 0.00042436,\ \ c_3 = 0.00001521. \tag{9}$$

This simple ELT window function, depicted in Fig. 3b, is less discontinuous at its borders than the proposals of [11, 16] and, as a result, allows the same level of side-lobe rejection as the double-length sine window of Fig. 3a. Concurrently, its main lobe remains narrower than that of the MLT sine window. Interestingly, it also resembles the latter window in shape.

To complete this section, Figure 3c illustrates the spectral and temporal shapes of the MDCT/MDST and ELT transition windows, based on the PC design of [24] and w_{elt} using (8, 9), and, for comparison, the double-length *start* window of AAC.

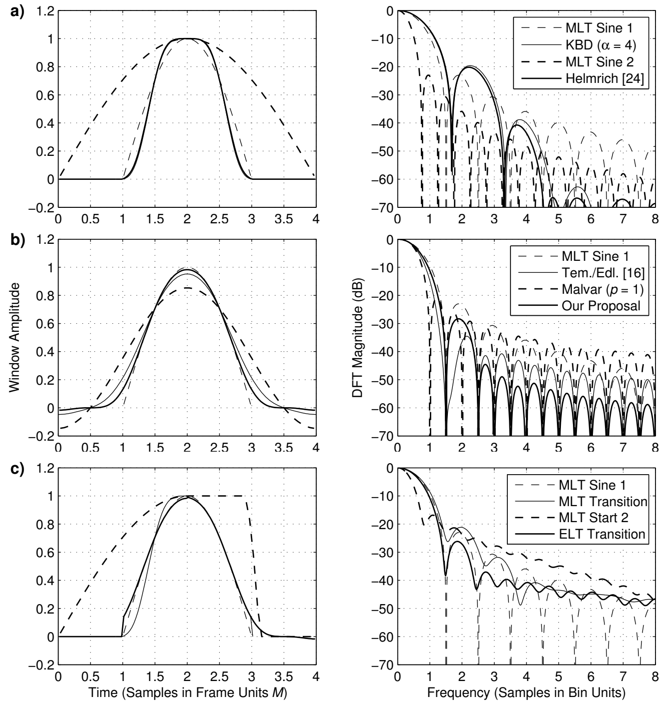

Fig. 3. *PR window designs:* (a) *MLT,* (b) *ELT,* (c) *transitions.*

4. INPUT-ADAPTIVE OVERLAP RATIO SELECTION

In order to verify its anticipated subjective advantage on tonal input and as a proof of concept, the switched coding approach of section 3.1, using the windows devised in section 3.2, may be integrated into a transform codec as follows. Note that, for reasons of brevity, only high-level aspects shall be discussed.

4.1. Decoder: Specification and Synthesis Transforms

An extra bit, signaling application of the ELT, is received per channel and/or frame in which long transformation (no block switching) has been utilized by the encoder. In case of MPEG coding the *window_shape* bit may be re-used for this purpose (1: MLT using window of [23] or [24], 0: our modified ELT). Based on this bit and the *window_sequence* (transform length and type), both for the current and last frame, the decoder can then deduce and apply the inverse lapped transform using the correct overlap ratio and window, as described in section 3.1.

4.2. Encoder: ELT Detector and Analysis Transforms

The encoder, applying and transmitting the per-channel/frame MLT/ELT choice such that encoder and decoder are synchronized, can detect stationary harmonic frames by computing an order-16 linear-prediction residual of the $\frac{1}{2}$-rate downsampled input, as done in speech coders [25], and deriving therefrom

- *temporal flatness* f_t as the ratio between the next and current frame's residual energy, with stationarity specified as $f_t < \frac{55}{8}$,
- *spectral flatness* f_s, also known as Wiener entropy, obtained from the DFT power spectrum of the current and next frame's concatenated residual, with high tonality indicated by $f_s < \frac{3}{8}$.

Figure 4 depicts the resulting frame-wise ELT (0) and MDCT (–1) selection for five input items (MDSTs are not used on this material). The stationary, tonal passages are detected reliably.

5. CODEC INTEGRATION AND EVALUATION

Two blind listening experiments according to the MUSHRA (*mu*ltiple *s*timuli with *h*idden *r*eference and *a*nchor) principle [26] were conducted to assess the subjective performance of the switched MDCT–ELT coding system in comparison with a conventional scheme employing only MDCTs (or MDSTs, as in case of the kernel switching proposal [9]). To this end, the switched-ratio architecture was integrated into an encoder and decoder implementation of the MPEG-H 3D Audio codec, using IGF for bandwidth extension and stereo filling (SF) for semi-parametric channel-pair coding at 48 kbit/s stereo, as described in [8, 9]. Testing was carried out by 12 experienced listeners (age 39 and younger, incl. 1 female) in a quiet room using a fanless computer and modern STAX headphones.

5.1. 48-kbit/s Test using Tonal Instrumental Signals

The first experiment intends to quantify the advantage of ELT over traditional MDCT coding on tonal, harmonic audio material, as well as the benefit of switching from ELT to MDCT coding on transients and tone onsets, as discussed in the last section. For each of the four tonal test signals already used in past MPEG codec evaluations [25, 27] – accordion, bag/pitch pipe, and harpsichord – the 3D Audio coded stimuli with and without the switchable ELT were presented alongside a 3D Audio reference condition employing *unified stereo* SBR and MPEG Surround 2-1-2 (and, thus, doubled frame length).

The results of this test, along with the per-stimulus 95-% confidence intervals, are illustrated as overall mean scores in Figure 5a and as differential mean scores, relative to the ELT condition, in Figure 5b. They demonstrate that for three out of the four items, the quality of the SBS-based 3D Audio codec is improved significantly by switching to the modified ELT in case of stationary passages. Moreover, by resorting to MDCT coding during non-stationarities, perceptual degradations due to stronger pre-echo artifacts are avoided. Finally, the performance of the SBS configuration is brought closer to that of the longer-frame-size *unified stereo* reference for such items; although block switching to 8 short transforms has not yet been implemented, all stimuli except *sm01* (bag pipes) now exhibit *good* quality.

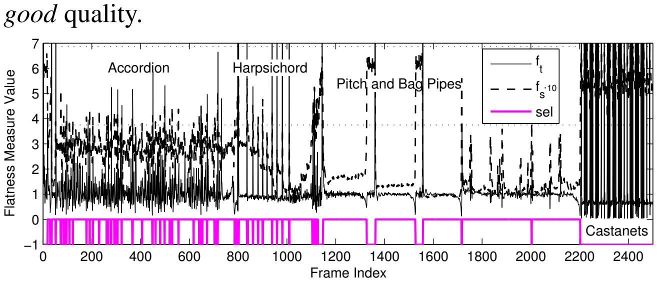

Fig. 4. *Spectral and temporal flatness based selection of ELT.*

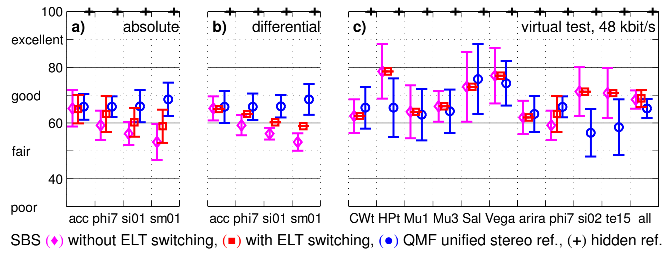

Fig. 5. *Zoomed view of listening test results with 95% confidence intervals. 3.5-kHz anchor scores are omitted for clarity.*

5.2. 48-kbit/s Virtual Test using Various Signal Types

A second "virtual" listening test was constructed in which the results of the subjective evaluation in [9] were combined with the present data for the *phi7* item (pitch pipe, the only signal in [9] for which ELTs are applied in more than a few frames). This setup shall reveal whether SBS-based 3D Audio coding, enhanced by the switchable ELT scheme, can outperform the QMF-based 3D Audio configuration on a diverse test set.

Figure 5c depicts the per-stimulus and the overall absolute mean scores, again with confidence intervals, for this test. Indeed, thanks to the ELT-induced quality gains on signals such as *phi7*, the average perceptual performance of the SBS+ELT configuration is rendered significantly better than that of the *unified stereo* reference. Given that the latter exhibits a higher algorithmic latency and complexity due to the required additional pseudo-QMF banks, this outcome is highly satisfactory.

6. CONCLUSION

This paper revisited the idea of extending the overlap ratio in lapped transform coding to more than the conventional 50% employed in contemporary audio codecs. The formulation of the resulting extended lapped transform (ELT) was modified to allow perfectly reconstructing transitions, with proper time domain aliasing cancelation, between MDCT coding with an overlap ratio of 50 and ELT coding with a ratio of 75%. A new ELT window with low side-lobe levels and, thus, less framing artifacts in typical applications was further designed. Finally, a complete system based on the MPEG-H 3D Audio specification [6, 7] was proposed, which can switch input adaptively between MDCT, MDST, and cosine- or sine-modulated ELT coding while allowing short-transform coding of transients.

The perceptual benefit of the switched-ELT approach was confirmed by formal subjective evaluation, which reveals no quality degradations over the 3D Audio framework and which further indicates that the authors' long-term objective of *good* coding quality on every type of input signal at 48 kbit/s stereo could in fact be achieved with only a bit more encoder tuning and a full implementation of ELT coding and block switching.

7. ACKNOWLEDGMENT

The authors thank the participants of the listening experiment.

8. REFERENCES

[1] J. P. Princen and A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1153–1161, Oct. 1986.

[2] J. P. Princen, A. W. Johnson, and A. B. Bradley, "Subband/transform coding using filter bank design based on time domain aliasing cancellation," in *Proc. of IEEE ICASSP '87*, Apr. 1987, vol. 12, pp. 2161–2164.

[3] H. S. Malvar, "Lapped Transforms for Efficient Transform/Subband Coding," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 38, no. 6, pp. 969–978, June 1990.

[4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, Oct. 1997.

[5] ISO/IEC MPEG-2 13818-7, "Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)," Dec. 1997.

[6] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio – The New Standard for Universal Spatial/3D Audio Coding," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 821–830, Dec. 2014.

[7] ISO/IEC MPEG-H 23008-3, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio," Mar. 2015.

[8] C. R. Helmrich, A. Niedermeier, S. Bayer, and B. Edler, "Low-complexity semi-parametric joint-stereo audio transform coding," in *Proc. of EUSIPCO '15*, Sep. 2015.

[9] C. R. Helmrich and B. Edler, "Signal-adaptive transform kernel switching for stereo audio coding," in *Proc. of IEEE WASPAA '15*, New Paltz, Oct. 2015, pp. 1–5.

[10] C. R. Helmrich, A. Niedermeier, S. Disch, and F. Ghido, "Spectral envelope reconstruction via IGF for audio transform coding," in *Proc. of IEEE ICASSP '15*, Apr. 2015, pp. 389–393.

[11] H. S. Malvar, "Modulated QMF Filter Banks with Perfect Reconstruction," *Electronics Letters*, vol. 26, no. 13, pp. 906–907, June 1990.

[12] H. S. Malvar, "Extended Lapped Transforms: Properties, Applications, and Fast Algorithms," *IEEE Trans. Signal Proc.*, vol. 40, no. 11, pp. 2703–2714, Nov. 1992.

[13] R. L. de Queiroz and K. R. Rao, "Adaptive extended lapped transforms," in *Proc. of IEEE ICASSP '93*, Apr. 1993, vol. 3, pp. 217–220.

[14] R. L. de Queiroz and K. R. Rao, "Time-Varying Lapped Transforms and Wavelet Packets," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3293–3305, Dec. 1993.

[15] M. Temerinac and B. Edler, "LINC: A Common Theory of Transform and Subband Coding," *IEEE Trans. Communications*, vol. 41, no. 2, pp. 266–274, Feb. 1993.

[16] M. Temerinac and B. Edler, "Overlapping Block Transform: Window Design, Fast Algorithm, and an Image Coding Experiment," *IEEE Trans. Communic.*, vol. 43, no. 9, pp. 2417–2425, Sep. 1995.

[17] G. D. T. Schuller and T. Karp, "Modulated Filter Banks with Arbitrary System Delay: Efficient Implementations and the Time-Varying Case," *IEEE Trans. Signal Proc.*, vol. 48, no. 3, pp. 737–748, Mar. 2000.

[18] H. S. Malvar, "A modulated complex lapped transform and its applications to audio processing," in *Proc. of IEEE ICASSP '99*, Mar. 1999, vol. 3, pp. 1421–1424.

[19] B. Edler, *Äquivalenz von Transformation und Teilbandzerlegung (Subband Decomposition) in der Quellencodierung*, Ph.D. thesis, Univ. Hannover, Germany, 1995.

[20] S. Shlien, "The Modulated Lapped Transform, Its Time-Varying Forms, and Its Applications to Audio Coding Standards," *IEEE Trans. Speech and Audio Proc.*, vol. 5, no. 4, pp. 359–366, July 1997.

[21] M. Padmanabhan and K. Martin, "Some further results on modulated/extended lapped transforms," in *Proc. of IEEE ICASSP '92*, Mar. 1992, vol. 4, pp. 265–268.

[22] K. M. A. Hameed and E. Elias, "Extended lapped transforms with linear phase basis functions and perfect reconstruction," in *Proc. of IEEE ICECS '05*, Dec. 2005.

[23] L. D. Fielder, M. Bosi, G. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low-Complexity Transform-Based Audio Coding," *AES collected papers on Digital Audio Bit-Rate Reduction*, pp. 54–72, 1996.

[24] C. R. Helmrich, "On the Use of Sums of Sines in the Design of Signal Windows," in *Proc. of DAFx-10*, Graz, Sep. 2010, online at http://dafx10.iem.at/proceedings/.

[25] M. Neuendorf, M. Multrus, N. Rettelbach, G. Fuchs, J. Robilliard, J. Lecomte, S. Wilde, S. Bayer, S. Disch, C. R. Helmrich, R. Lefebvre, P. Gournay, B. Bessette, J. Lapierre, K. Kjörling, H. Purnhagen, L. Villemoes, W. Oomen, E. Schuijers, K. Kikuiri, T. Chinen, T. Norimatsu, K. S. Chong, E. Oh, M. Kim, S. Quackenbush, and B. Grill, "The ISO/MPEG Unified Speech and Audio Coding Standard – Consistent High Quality for all Content Types and at all Bit Rates," *J. Audio Eng. Soc.*, vol. 61, no. 12, pp. 956–977, Dec. 2013.

[26] ITU, Radiocommunication Sector, "Recommendation BS.1534-2: Method for the subjective assessment of intermediate quality level of audio systems," June 2014.

[27] K. Brandenburg and M. Bosi, "Overview of MPEG-Audio: Current and Future Standards for Low Bit-Rate Audio Coding," in *Proc. of AES 99th Convention*, New York, Oct. 1995, no. 4130.