Audio Coding Using Overlap and Kernel Adaptation

Christian R. Helmrich, *Graduate Student Member, IEEE,* and Bernd Edler

***Abstract*—Perceptual audio coding schemes typically apply the modified discrete cosine transform (MDCT) with different lengths and windows, and utilize signal-adaptive switching between these on a per-frame basis for best subjective performance. In previous papers, the authors demonstrated that further quality gains can be achieved for some input signals by using additional transform kernels such as the modified discrete sine transform (MDST) or greater inter-transform overlap by means of a modified extended lapped transform (MELT). This work discusses the algorithmic procedures and codec modifications necessary to combine all of the above features – transform length, window shape, transform kernel, and overlap ratio switching – into a flexible input-adaptive coding system. It is shown that, due to full time-domain aliasing cancelation, this system supports perfect signal reconstruction in the absence of quantization and, thanks to fast realizations of all transforms, increases the codec complexity only negligibly. The results of a 5.1 multichannel listening test are also reported.**

***Index Terms*—Audio coding, lapped transform, MDCT, MDST**

I. INTRODUCTION

PERCEPTUAL coding of digital audio signals by means of quantization and compression of spectral coefficients has relied on a type of lapped transformation known as modified discrete cosine transform (MDCT) for more than two decades. The MDCT, presented by Princen *et al.* in 1987 [1] and further examined by Malvar in 1990 under the name modulated lapped transform [2], represents an oddly stacked filter-bank design in which adjacent transforms are applied such that their temporal overlap is 50%. By utilizing low-overlap analysis and synthesis windows with the forward and inverse transforms, respectively, the inter-transform overlap can be minimized [3]. This reduces pre-echo artifacts (temporal smearing of the quantization error) when coding transient signal portions such as drum hits or tone onsets. Further pre-echo reduction can be attained using either input based adaptation of the instantaneous transform lengths, a technique commonly known as “block length switching” [4], or temporal noise shaping (TNS), a tool which applies forward adaptive linear predictive coding to certain MDCT coefficients [5]. The combination of block switching and TNS was shown to allow good coding quality even on very transient input such as castanet recordings [6] and has been adopted in all modern ISO/IEC audio standards, starting with MPEG-2 AAC [7] and culminating in the new MPEG-H 3D Audio codec [8], [9].

In [10], the authors demonstrated that for two-channel stereo input with an inter-channel phase difference (IPD) around ±90 degrees, considerable quality increase and decoder complexity reduction can be achieved by employing the type-IV modified discrete sine transform (MDST) instead of the MDCT in one of the channels. Furthermore, it was shown that the presented kernel switching scheme merely alters the transform definition without affecting the windowing and overlap-add (OLA) steps, thereby maintaining support for the window shape adaptation approach of [3]. Compatibility with the block length switching technique of [4] was also noted but not described in detail.

More recently, additional improvements for quasi-stationary harmonic signal passages by adaptive application of a modified extended lapped transform (MELT) were reported [11]. Being based on Malvar’s ELT formulation [12], the MELT constructs an oddly stacked filter-bank with 75% inter-transform overlap, as depicted in Figure 1b), yielding greater frequency selectivity than an MDCT or MDST filter-bank with 50% overlap, shown in Figure 1a), at the same frame length M. However, unlike the ELT, the MELT allows straight-forward transitions – requiring only special transitory windows – to and from MDCTs. In [11] a respective frame-wise signal-adaptive overlap ratio switching scheme was derived, and its benefit over ELT-only or MDCT-only coding was confirmed via subjective testing. As in [10] an integration into, and compatibility with, block length switching designs was not discussed; block switching functionality in the 3D Audio codec was disabled for the investigations in [11].

A. Contribution and Organization of this Article

In order to realize a fully flexible coding system, it is desirable to implement the following additional structural capability:

- kernel switching also in case of MELT coding for harmonic input with ±90 degrees of IPD (assuming synchronized MELT activation in the stereo channels). This is studied in section II.
- more flexibility and optimizations in the transitions from and to MELT coded frames. These aspects, motivated especially by a need for minimal encoding delay, are addressed in section III.
- block length switching in combination with kernel switching and / or MELT coding. Implementational details like the choice of transform sequences or windows are examined in section IV.

Section V discusses the preparation and outcome of a listening experiment evaluating the overlap- and kernel-switching codec proposal after incorporation of the enhancements presented in sections II–IV. Finally, section VI concludes the paper.

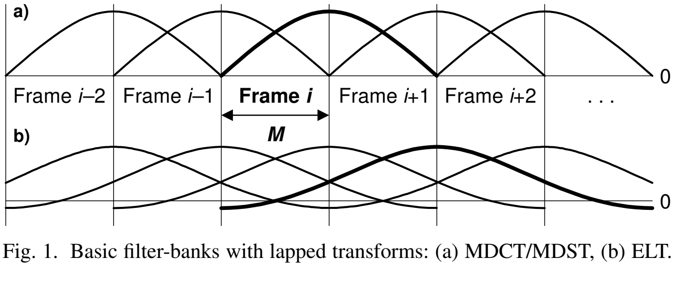

Fig. 1. Basic filter-banks with lapped transforms: (a) MDCT/MDST, (b) ELT.

Manuscript received December 22, 2015; revised January 27, 2016; accepted January 29, 2016. Date of publication February 8, 2016; date of current version February 9, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Panayiotis Georgiou.

The authors are with the International Audio Laboratories Erlangen, a joint institution of Fraunhofer IIS and Friedrich-Alexander University of Erlangen-Nürnberg, Germany (e-mail: christian.helmrich@audiolabs-erlangen.de).

II. COSINE- AND SINE-MODULATED MELT CODING

The forward (analysis) MDCT for a frame at index i, given a time signal x and returning a spectrum X, can be written as

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n)\, \cos\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr), \tag{1}$$

where window length $N = 2M$ and $0 \le k < M$. Likewise, the forward MDST is defined using a sine instead of cosine term:

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n)\, \sin\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr). \tag{2}$$

Altering the temporal length and phase offset yields the MELT,

$$X_i(k) = \sum_{n=0}^{L-1} x_i(n)\, \cos\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr) \tag{3}$$

with increased window length $L = 4M$ and cosine modulation. Naturally, a sine-modulated counterpart may also be specified,

$$X_i(k) = \sum_{n=0}^{L-1} x_i(n)\, \sin\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr), \tag{4}$$

as indicated in [11]. The inverse (synthesis) MELT variants are

$$x'_i(n) = \tfrac{2}{M} \sum_{k=0}^{M-1} X'_i(k)\, \cos\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr) \tag{5}$$

for cosine banks applying (3) and, respectively for sine banks,

$$x'_i(n) = \tfrac{2}{M} \sum_{k=0}^{M-1} X'_i(k)\, \sin\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + \tfrac{1}{2}\bigr)\Bigr), \tag{6}$$

where $'$ denotes optional spectral processing, e. g. quantization, and $0 \le n < L$. Notice that, even though the employed *window* length varies between (1, 2) and (3–6), the *transform* length M (and thereby, the inter-transform step size illustrated in Fig. 1) stays identical, which explains the difference in overlap ratio.

The cosine- and sine-modulated MELT definitions of (3–6) are not sufficient for realizing kernel switching – and therefore, efficient coding of signals with ±90 degrees of IPD – even in case of 75% inter-transform overlap. As shown in [10], type-II transition transforms adopted from the evenly stacked Princen-Bradley filter-bank [13] are necessary for time-domain aliasing cancelation (TDAC) when switching between type-IV MDCTs and MDSTs (1, 2). Specifically, a MDST-II is required during changes from MDCT-IV to MDST-IV coding in a channel, and a MDCT-II is needed when reverting to MDCT-IV coding.

Hameed and Elias [14] demonstrated that, beside the above-mentioned type-IV definitions of [12] and (3–6), an ELT-based filter-bank allowing fast implementations using the DCT-II can also be constructed, which proves that type-II filter-banks with more than 50% inter-transform overlap are in fact feasible. An alternative but equivalent approach following the TDAC filter-bank design [10, 13] is to devise an evenly stacked system via alternating usage of a type-II cosine-modulated MELT version,

$$X_i(k) = \tfrac{1}{1+\delta(k)} \sum_{n=0}^{L-1} x_i(n)\, \cos\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)k\Bigr) \tag{7}$$

with Kronecker delta $\delta(0) = 1$, and a type-II sine-based MELT,

$$X_i(k) = \tfrac{1}{1+\delta(k')} \sum_{n=0}^{L-1} x_i(n)\, \sin\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + 1\bigr)\Bigr) \tag{8}$$

with $k' = M-1-k$ for scaling of the Nyquist coefficient as in [10]. Proving that (7, 8) on the analysis side and, respectively,

$$x'_i(n) = \tfrac{2}{M} \sum_{k=0}^{M-1} X'_i(k)\, \cos\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)k\Bigr) \tag{9}$$

and

$$x'_i(n) = \tfrac{2}{M} \sum_{k=0}^{M-1} X'_i(k)\, \sin\Bigl(\tfrac{\pi}{M}\bigl(n + \tfrac{3M+1}{2}\bigr)\bigl(k + 1\bigr)\Bigr) \tag{10}$$

on the synthesis side lead to TDAC, as depicted in Figure 2, is straight-forward and will be omitted for the sake of brevity.

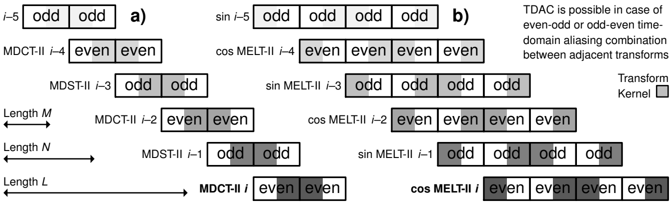

Fig. 2. TDAC in evenly stacked filter-banks: (a) Princen-Bradley, (b) MELT-II.

Unfortunately, regarding the combination of MELT coding and kernel switching, it can be shown that TDAC is impossible when, analogously to the process for 50% overlap, a transitory type-II instance of (7, 9) or (8, 10) is employed when switching between type-IV cosine- and sine-modulated MELTs (3–6). As it is desirable to keep the architectural complexity of the codec low when allowing kernel switching regardless of the instantaneous overlap ratio, the following work-around is proposed.

To switch from the cosine-modulated MELT-IV (3, 5) to the sine-modulated MELT-IV (4, 6) a transitory MDST-II frame as in [10], combined with a necessary temporary reduction of the overlap ratio to 50% on both analysis and synthesis side, can be used. Likewise, an intermediate MDCT-II can be employed when reverting back from sine- to cosine-based MELT coding. Full TDAC is guaranteed in both cases since, as is visualized in Figure 3, the overlap length between each type-II transition and its type-IV MELT neighbors is restricted to $M = \frac{N}{2}$ (hence, there is no temporal-aliasing bound overlap between a cosine- and a sine-modulated MELT-IV which requires TDAC). Proper windowing, however, is crucial – a special “stop-start” window must be applied to the type-II transforms, as shown in Figure 4 a). This symmetric window, which is based on the asymmetric transitory weightings derived in [11], is described in Sec. III.

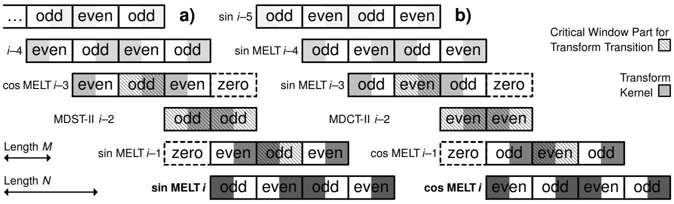

Fig. 3. Proposed TDAC-compliant kernel switching for MELT-IV filter-banks: transitions from (a) cosine to sine modulation, (b) sine to cosine modulation.

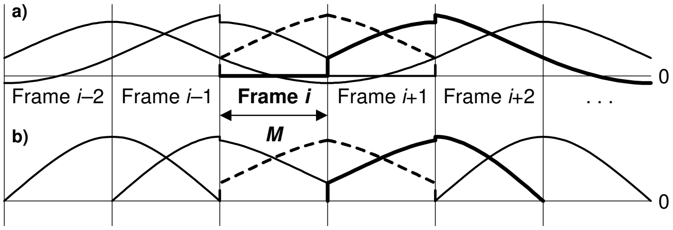

Fig. 4. Correct windowing with (- - -) special "stop-start" shape during temporary transitions from (a) 75 to 50% overlap ratio, (b) 50 to 75% overlap ratio.

III. TRANSITIONS FROM AND TO MELT FRAMES

In [11] the authors introduced a solution for frame-to-frame switches from an MDCT-like transform with 50% to the MELT with 75% overlap ratio, and vice versa. To maintain full TDAC during the switches, dedicated asymmetric transition windows derived from the steady-state weightings applied during quasi-stationary signal passages were proposed. These are defined as

$$w'_{\text{elt}}(n) = \begin{cases} 0, & 0 \le n < M \\ w_{\text{elt}}(n), & M \le n < N \\ d\sqrt{1 - w_{\text{elt}}(k)^2 - w_{\text{elt}}(M+k)^2}, & N \le n < 3M \\ w_{\text{elt}}(n), & 3M \le n < L \end{cases} \tag{11}$$

for the first MELT window upon an overlap increase from 50 to 75% (bold-lined shape depicted in Fig. 4a) for frame i) and

$$w'_{\text{mlt}}(n) = \begin{cases} d\sqrt{1 - w_{\text{elt}}(N+k)^2 - w_{\text{elt}}(3M+k)^2}, & 0 \le n < M \\ w_{\text{mlt}}(n), & M \le n < N \end{cases}$$

for the first MDCT/MDST window when reducing the overlap ratio to 50% (bold-lined shape in Fig. 4b) for the same frame). Parameter d, which equals 1 in [11], is examined below. The complements for w'_{elt} and w'_{mlt} – the last MELT window when switching to 50% overlap, and the last MDCT/MDST window during switch-backs to 75% overlap (frame $i-2$ in Fig. 4) – are simply the temporal reversals of w'_{elt} and w'_{mlt}, respectively. k, used in the critical window parts (see also Fig. 3), is specified as above, while w_{elt} resp. w_{mlt} indicate the underlying window functions for a steady-state MELT and MDCT/MDST. For the former, which is also applicable to the ELT [12], an improved design preventing blocking artifacts was devised in [11]. This window is also used herein for the following investigations.

Using $d \ne 1$ allows for a generalized biorthogonal approach regarding switched-ratio transition windowing, where different critical window parts, denoted w_{tr} hereafter, may be employed for the analysis and synthesis transforms. Specifically, to reach TDAC, w'_{elt} and w'_{mlt} may use $d = d'$ on the analysis (encoder) side and $d = \frac{1}{d'}$ on the synthesis (decoder) side. Given a steady-state w_{elt}, d in w_{tr} is preferably selected such that, during ratio switching transitions, the following attributes are obtained:

- optimal spectral properties of the analysis windows to maximize the coding gain especially on stationary harmonic signals. Given optimized steady-state weightings, this can be achieved by eliminating the discontinuities at the center of the transitory windows. Specifically, d' shall be a ratio chosen such that the maximum of w'_{elt} equals the maximum of w_{elt}, here $d' = \frac{4096}{4061}$.
- maximum output attenuation upon synthesis windowing as a means to suppress distortion caused by the spectral processing (symbolized by $'$ in Sec. II) as much as possible prior to OLA completion. To prevent rendering the complementary analysis windows detrimental with regard to efficiency, $d \ll 1$ should be avoided. A good trade-off can be achieved using $\frac{1}{d'} = \frac{4061}{4096}$.

In other words, both optimization approaches for w_{tr} preferably result in the same value for d'. Note, however, that when using w_{elt} from [11], the center discontinuities in w'_{elt} and w'_{mlt} are very minor (see Fig. 5), and their avoidance at least on the analysis side is not expected to yield audible improvement. In fact, the *border* discontinuities degrade the audio quality much more. w'_{elt} and w'_{mlt} should generally be applied only rarely.

In Sec. II a special transitory "stop-start" window for MELT-based kernel switching was introduced. This window, depicted by a dashed line in Fig. 4a) and denoted by w_{ss} hereafter, can be easily derived from the critical window part of w'_{elt} or w'_{mlt}:

$$w_{\text{ss}}(n) = \begin{cases} d\sqrt{1 - w_{\text{elt}}(N+k)^2 - w_{\text{elt}}(3M+k)^2}, & 0 \le n < M \\ d\sqrt{1 - w_{\text{elt}}(k)^2 - w_{\text{elt}}(M+k)^2}, & M \le n < N. \end{cases}$$

More specifically, w_{ss} is a symmetric window with critical parts in both halves, thus allowing overlap-ratio transitions on both sides. It is worth noting that w_{ss} can be applied to the MDCT and MDST as well as the different MELT variants (assuming the outer quarters of the length-L weighting are set to zero). In fact, its usage for analysis-side windowing renders the MDCT and the cosine-modulated MELT-IV coefficients identical apart from sign differences, as is indicated by Figure 2c) in [11].

Besides facilitating kernel switching, w_{ss} can also be utilized to make the overlap ratio adaptation scheme more flexible. For instance, a temporary switching configuration from 50 to 75% overlap, as shown in Fig. 4b), can be achieved therewith. Such a short-term overlap increase is useful when an objective is to minimize the encoder-side look-ahead used for ratio detection. To provide an example, assume that a total encoder look-ahead of $L-M = 3M$ shall not be exceeded. As can be observed in Fig. 4b), if the ratio detector chooses a switch-back to MELT coding in frame $i-2$, e. g. after a transient (thin-lined transitory asymmetric MDCT/MDST window), but in the next frame $i-1$ detects a new transient at the end of the look-ahead region, i. e. the length-M segment labeled "Frame $i+2$", it must instantly revert to 50%-ratio coding to minimize the inevitable pre-echo [11]. Since the first half of the MELT window for $i-1$ (dashed line) has already been determined by the complementary shape for $i-2$ and, to preserve TDAC, cannot be changed anymore, only the second window half can be freely selected. Mirroring the first to the second half produces the TDAC compliant w_{ss} shape depicted in Fig. 4b), which can be followed by the w'_{mlt} shape (bold line) in i. Note, however, that this solution exhibits notable window-border discontinuities at three places – frames $i-1$, i, and $i+1$ – leading to reduced spectral compaction and, most likely, suboptimal coding quality over this period. It thus

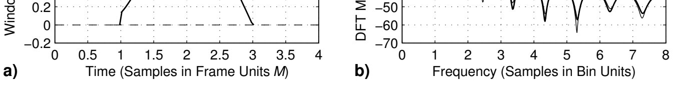

Fig. 5. Effect of removing discontinuity at center of the transitory w'_{mlt} shape: (a) values of window weights, (b) magnitudes of respective transfer functions.

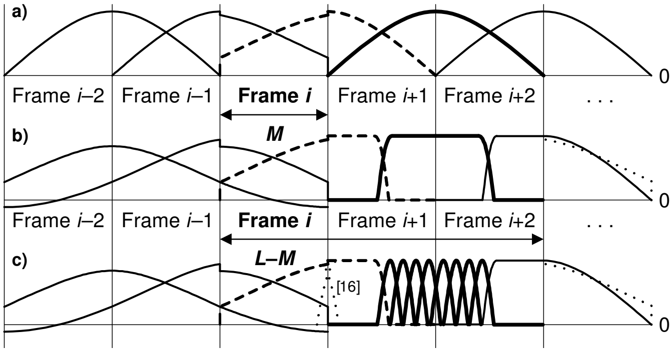

Fig. 6. Combination of MELT and block length switching. See text for details.

seems preferable to, as an alternative, avoid the symmetric w_{ss} in this case and utilize the asymmetric w'_{mlt} already in $i{-}1$. In doing so, the steady-state w_{mlt} can be applied in i, as shown in Figure 6a), and the border discontinuity at $i{+}1$ can be avoided.

IV. Block Switching, Kernel Switching, and MELT

It is well-known that the analysis and synthesis windows of a MDCT-based filter-bank can be adapted to the instantaneous input characteristics on a per-transform basis without violating the TDAC principle [3], [4]. Naturally, the same holds for our kernel switching approach since, as noted in Sec. I, transitions between MDCTs and MDSTs do not affect the windowing and OLA processes [10]. Consequently, switching to a low-overlap window shape during transient signal passages is also possible while kernel switching is being utilized (assuming valid transform sequences are still applied, as discussed in Sec. II) and/or in case of MELT coding (assuming the overlap ratio has been reduced to 50% in the preceding frame, as proposed in Sec. II for a kernel change). Fig. 6b) depicts this window sequence.

The length-N dashed window shape for $i{-}1$ in Fig. 6b) is a concatenation of the first half of w'_{mlt} or w_{ss} and the second half of the "long start" window used since MPEG-2 AAC [15]. Accordingly, the first half of the MDCT/MDST window for i must equal the first M samples of the AAC "long stop" shape. The latter is illustrated for frame $i{+}1$ to indicate a switch-back to full 50% of overlap ratio (and possibly 75%, note the dotted line). It is easy to see that block length switching can now be realized by simply replacing the single low-overlap transform at i by eight consecutive 50% overlapped short transforms, as in AAC. The resulting configuration is visualized in Fig. 6c), including a low delay block switch via transform splitting [16].

V. Objective and Subjective Evaluation

To assess the performance of our consolidated flexible codec proposal – allowing simultaneous window shape, block length, kernel, and overlap ratio switching – the extensions described in the previous two sections were integrated into Fraunhofer's implementation of the 3D Audio core-codec, which had been enhanced already by the fundamental kernel and overlap ratio adaptation schemes [10], [11]. The encoder-side window shape and block length detectors operated as usual, and their outputs (i. e. signal metadata) were used as control inputs to the kernel and MELT detection. Both objective (complexity analysis) and subjective testing (formal blind listening experiment evaluating the basic/overall audio quality) was carried out.

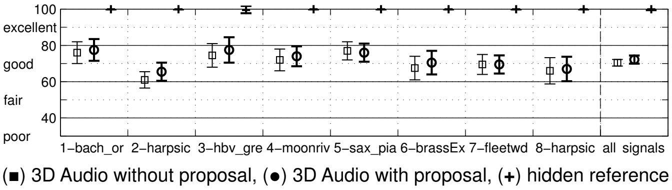

Fig. 7. Zoomed view of listening test results. 3.5-kHz anchor scores not shown.

A. Assessment of Computational Complexity Increase

The cosine- and sine-modulated MELT-IV defined by (3, 5) and (4, 6) were implemented by way of MDST-IV and MDCT-IV, respectively, assisted by additional sign flips for the transform coefficients at the even or odd indices and L-to-N (analysis) or N-to-L (synthesis) folding operations. Given that these N-size MDST-IV and MDCT-IV, in turn, can be realized via discrete cosine or sine transforms of length M [2], [10] like the other 50%-overlap transforms employed in the codec and, hence, do not add any complexity, only the extra sign flips and foldings as well as doubled window and OLA lengths must be considered. The latter result in roughly $3L$ additional operations per frame, which is only about 2% of the total codec complexity.

B. Perceptual Evaluation at 80kbps 5.1 Multichannel

As in [10], [11] a blind subjective test following the MUSHRA (*mul*tiple *s*timuli with *h*idden *r*eference and *a*nchor) paradigm [17] was conducted. Both 3D Audio variants tested – with and without unified MELT and kernel switching – were operated at a bandwidth of 16 and a sample rate of 32 kHz (which was upsampled to the input rate of 48 kHz for listening) with a bitrate of 80 kbit per second for a typical 5.1 channel setup. For the LFE channel both kernel and overlap ratio switching were deactivated, and the last seven transforms of each "eight short" sequence were restricted to the type-IV MDCT or MDST (depending on the first transform, whose type was unconstrained). The remaining codec parameters were configured as described in [10], [11]. 10 experienced listeners (age 38 or younger, incl. one female) performed the experiment in a quiet room using a silent computer and modern Dynaudio BM6A loudspeakers.

Figure 7 illustrates the results of the listening test as overall and per-stimulus mean scores along with their associated 95% confidence intervals. For 6 of the 8 signals tested (tonal and/or transient recordings from past EBU multichannel tests [18]) an audio quality increase can be attributed to the inclusion of the enhanced MELT and kernel switching. Further analysis reveals that for two items (*2-harpsichord* and *6-brassEx*) as well as the overall mean score, the improvement is statistically significant.

VI. Conclusion

This article unified and extended the authors' prior work on kernel and overlap ratio switching for flexible use in block size adapting audio coders by studying the required changes to the transform sequencing and windowing. Both full TDAC and a low codec complexity could be maintained, and small but significant gains in coding quality were achieved on some signals.

Acknowledgment

The authors thank F. Schuh and T. Schwegler for their help in the implementation and subjective evaluation of the codec.

REFERENCES

[1] 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. IEEE ICASSP ’87*, Dallas, vol. 12, pp. 2161–2164, Apr. 1987.

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

[3] E. Allamanche, R. Geiger, J. Herre, and T. Sporer, “MPEG-4 Low Delay Audio Coding based on the AAC Codec,” in *Proc. AES 106th Convention*, Munich, preprint 4929, May 1999.

[4] B. Edler, “Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen (Coding of audio signals with overlapped transform and adaptive window functions),” *Frequenz*, v. 43, 1989.

[5] J. Herre and J. D. Johnston, “Continuously signal-adaptive filterbank for high-quality perceptual audio coding,” in *Proc. IEEE WASPAA*, Oct. 1997.

[6] C. R. Helmrich, G. Marković, and B. Edler, “Improved low-delay MDCT-based coding of both stationary and transient audio signals,” in *Proc. IEEE ICASSP ’14*, Florence, pp. 6954–6958, May 2014.

[7] ISO/IEC 13818-7, “Information technology – Generic coding of moving pictures and associated audio information – Part 7: AAC,” Dec. 1997.

[8] ISO/IEC 23008-3, “Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio,” 2015.

[9] ISO/IEC SC29/WG11, N15849, “Text of ISO/IEC 23008-3:201x/DAM 3, MPEG-H 3D Audio Phase 2,” Nov. 2015.

[10] C. R. Helmrich and B. Edler, “Signal-adaptive transform kernel switching for stereo audio coding,” in *Proc. IEEE WASPAA*, pp. 91–95, Oct. 2015.

[11] C. R. Helmrich and B. Edler, “Signal-adaptive switching of overlap ratio in audio transform coding,” in *Proc. IEEE ICASSP*, Shanghai, Mar. 2016.

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

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

[14] K. M. A. Hameed and E. Elias, “Extended lapped transforms with linear phase basis functions and perfect reconstruction,” in *Proc. IEEE ICECS ’05*, Gammarth, pp. 1–4, Dec. 2005.

[15] M.Bosi, K.Brandenburg, S.Quackenbush, L.Fielder, K.Akagiri, M.Dietz, J. Herre, G. Davidson, and Y. Oikawa, “ISO/IEC MPEG-2 Advanced Audio Coding,” *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[16] C. R. Helmrich and M. Fischer, “Low-Delay Transform Coding Using the MPEG-H 3D Audio Codec,” in *Proc. AES 139th Convention*, New York, preprint 9355, Oct. 2015.

[17] ITU, radiocommunication sector, “Recommendation BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems,” Oct. 2015, online at www.itu.int/rec/R-REC-BS.1534-3-201510-I.

[18] EBU, technical report 3324, “EBU Evaluations of Multichannel Audio Codecs,” Sep. 2007, online at https://tech.ebu.ch/docs/tech/tech3324.pdf.