Non-Uniform Orthogonal Filterbanks Based on MDCT Analysis/Synthesis and Time Domain Aliasing Reduction

Nils Werner and Bernd Edler

***Abstract*—In this paper we describe non-uniform orthogonal modified discrete cosine transform (MDCT) filterbanks and time domain aliasing reduction (TDAR). By adding a post-processing step to the MDCT, our method allows for arbitrary non-uniform frequency resolutions using subband merging with smooth windowing and overlap in frequency. This overlap allows for an improved temporal compactness of the impulse response, which is especially useful for audio coders. The post-processing step comprises another lapped MDCT transform along the frequency axis and TDAR along each subband signal.**

***Index Terms*—TDAC, MDCT, Perceptual Coding, Time-Frequency Transform**

I. Introduction

THE MDCT is widely used in audio coding applications due to properties such as good energy compaction and critical sampling when used with appropriately shaped overlapping windows.

In perceptual audio coding, quantization noise introduced by the coder needs to be both spectrally and temporally shaped such that it does not exceed the masking threshold. The MDCT exhibits a uniform time-frequency resolution but the masking threshold has a non-unifom spread in time and frequency [1]. Consequently, a non-uniform time-frequency resolution is a desirable representation. Such a non-uniform transform allows individual temporal and spectral shaping of the quantization noise in different bands and allows to more closely match the non-uniform perception characteristics of the human ear [2]. This leads to a higher coding efficiency when compared to uniform filterbanks [3]. Critically, to allow precise shaping of the quantization noise and thus be suitable for coding, the resulting filterbank is required to have both temporally and spectrally compact impulse responses [3].

A non-uniform filterbank based on Hadamard merging [4] is implemented in a current audio codec [5]. However, it is rarely activated as the poor temporal compactness of the method severely limits the maximum possible non-uniformity and coding efficiency [3].

A way of designing non-uniform FIR filterbanks is the repeated application of one or more uniform transforms [6]:

N. Werner and B. Edler are with the International Audio Laboratories Erlangen, a joint research institute between the Friedrich-Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer Institute of Integrated Circuits, IIS, Erlangen, Germany (e-mail: nils.werner@audiolabs-erlangen.de)

For *subband merging*, a long transform is applied first, transforming the signal from the temporal to the spectral domain. The result is a spectrum with high spectral but low temporal resolution. Afterwards, selected groups of spectral bins (a *subband*) are transformed to a subband signal. This increases the temporal resolution while sacrificing spectral resolution in that selected subband.

Subband splitting is the complementary operation: First a short transform is applied. The result is a spectrum with low spectral but high temporal resolution. Afterwards, the spectral bins of two or more adjacent transform frames are transformed again, increasing their spectral resolution at the cost of temporal resolution.

These steps and their sizes can be mixed and repeated at will. The choice of transform can be arbitrary, however, the same or a similar transform for each step is expected to have the best results [6].

There are numerous ways of facilitating non-uniform time-frequency transforms using repeated uniform transforms [7], e.g. using two successive Fourier transforms [8], discrete cosine transform type IV (DCT-IV) followed by MDCT subband merging (MSM) [9], MDCT followed by butterfly element merging [10] and MDCT followed by Hadamard subband merging (HSM) [4].

Of these transforms, HSM is the only one that allows direct integration into common audio coding pipelines. But HSM allows only for limited frequency scale designs with subband widths constrained to powers of two. The subband merging operation is performed without any overlap, which results in compact frequency responses but very long impulse responses [4], [5]. Also the Hadamard matrix only very roughly approximates the DCT and thus allows for only limited overall spectrotemporal resolution. For an analysis see [3] and Section IV.

Additionally, despite using the MDCT, which produces time domain aliasing, HSM does not reduce the aliasing in the subbands, resulting in a long filterbank impulse response with low temporal compactness. A possible, aliasing reducing subband merging solution would be to use a transform which is followed by time domain aliasing cancellation (TDAC), like IMDCT.

This paper presents a new method which works on a sequence of MDCT spectra and uses direct MDCT as the subband merging operation, followed by time domain aliasing reduction (TDAR). The resulting nonuniform filterbank is lapped in both time and frequency, orthogonal and allows for

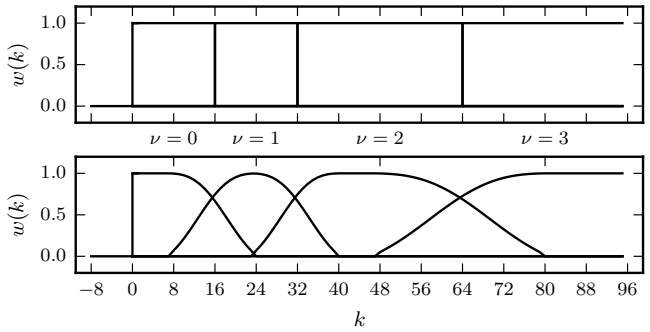

Fig. 1. Example subband layouts using IMDCT Subband Merging (top, rectangular windows without spectral overlap) and MDCT Subband Merging (bottom, Kaiser Bessel derived windows with spectral overlap). Note the first window quarter being zero and overlapping beyond zero at $\nu = 0$.

any even subband widths. Due to overlap along the frequency axis and TDAR, a temporally highly compact subband impulse response is achieved.

II. Filterbank Implementation

A. Analysis Filterbank

The analysis filterbank implementation directly builds upon the common lapped MDCT; the original transform with overlap and windowing remains unchanged.

$$x_i(n) = x(n + iM) \quad 0 \le n < 2M \tag{1}$$

$$X_i(k) = \sqrt{\frac{2}{M}} \sum_{n=0}^{2M-1} h(n) x_i(n) \kappa(k, n, M) \quad 0 \le k < M \tag{2}$$

where i is the frame index and $h(n)$ is a suitable analysis window. To make notation easier, $\kappa(k, n, M)$ is a slightly modified, but also occasionally used MDCT transform kernel [11], [12] with $n_o = -M/2+1/2$ for n_0 as introduced in [13], [14]

$$\kappa(k, n, M) = \cos\left[\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(n - \frac{M}{2} + \frac{1}{2}\right)\right]. \tag{3}$$

The standard MDCT transform kernel with $n_o = M/2 + 1/2$ can also be used, in which case time-reversal in the resulting subband signals has to be taken into account [15].

We recall that the direct MDCT can be interpreted as windowing, pre-permutation and DCT-IV. The inverse MDCT, can be interpreted as DCT-IV, post-permutation, windowing and overlap-add, called TDAC [16], [17]. Note that the DCT-IV, which is its own inverse, is used in both direct and inverse transform.

For IMDCT subband merging, the spectrum $X_i(k)$ is segmented into subbands of individual widths N_ν and each subband is transformed back to subband time domain using IMDCT of length N_ν. However, due to its rectangular spectral windowing without overlap, the IMDCT is expected to have high temporal ringing and thus inhibit compact impulse responses. By replacing the IMDCT with a lapped direct MDCT of window length $2N_\nu$, we can relax the brick-wall rectangular subband edges with a smooth analysis window and overlap and reduce the undesired ringing. This method has also been used in [9] on DCT-IV spectra. However, since the direct MDCT does not have a following TDAC stage, the resulting subband samples will contain time domain aliasing from (2). A method that combines TDAC and approximate window coefficients to reduce the aliasing will later be introduced in Section III. Also, we have now introduced aliasing in the frequency domain, which must be cancelled in synthesis.

For the sake of brevity we are using one common merge factor $N = N_\nu$ for all subbands ν, however any valid MDCT window switching sequence can be employed to create the desired subband widths [18]. More resolution design considerations will be introduced in Section II-C. The subband merging operation then results in

$$X_{\nu,i}(k) = X_i\left(k + \nu N\right) \quad 0 \le k < 2N \tag{4}$$

$$\hat{y}_{\nu,i}(m) = \sqrt{\frac{2}{N}} \sum_{k=0}^{2N-1} w(k) X_{\nu,i}(k) \kappa(m, k, N) \quad 0 \le m < N \tag{5}$$

where ν is the subband index, and $w(k)$ is a suitable analysis window that generally differs from $h(n)$ in size and may differ in window type.

The output $\hat{y}_{\nu,i}(m)$ is a set of vectors of individual lengths of N_ν subband samples.

B. Synthesis Filterbank

For synthesis, the inner MDCT (5) is inverted using the IMDCT, the synthesis window $v(k)$, and TDAC (albeit the aliasing cancellation is done along the frequency axis k), perfectly reconstructing the spectrum $X_i(k)$:

$$\hat{X}_{\nu,i}(k) = \sqrt{\frac{2}{N}} \sum_{m=0}^{N-1} \hat{y}_{\nu,i}(m) \kappa(k, m, N) \quad 0 \le k < 2N \tag{6}$$

$$X_{\nu,i}(k) = v(k + N)\hat{X}_{\nu-1,i}(k + N) + v(k)\hat{X}_{\nu,i}(k) \tag{7}$$

$$X_i\left(k + \nu N\right) = X_{\nu,i}(k)\,. \tag{8}$$

Afterwards, the outer MDCT (2) is inverted using the IMDCT, the synthesis window $g(n)$, and TDAC (here aliasing cancellation is done along the time axis n), perfectly reconstructing the time domain signal.

$$\hat{x}_i(n) = \sqrt{\frac{2}{M}} \sum_{k=0}^{M-1} X_i(k) \kappa(n, k, M) \quad 0 \le n < 2M \tag{9}$$

$$x_i(n) = g(n + M)\hat{x}_{i-1}(n + M) + g(n)\hat{x}_i(n) \tag{10}$$

$$x(n + iM) = x_i(n)\,. \tag{11}$$

C. Time-Frequency Resolution Design Limitations

While any desired time-frequency resolution is possible, some constraints for designing the resulting window functions must be considered to ensure invertibility. In particular, the squares of two adjacent TDAR window slopes must add to one so that (5), (7) fulfill the Princen Bradley condition [14]. The window switching scheme as introduced in [18], originally designed to reduce pre-echo effects, can be applied here [9].

For correct border handling, (4)–(8) must be shifted by $-N_0/2$ so that the first quarter of the $\nu = 0$ window (the first aliasing term) extends beyond zero, combined with rectangular start window half at the border (Fig. 1) [19]. The same must be done for the last band. Bands may be excluded from merging using this technique with zeros at the desired coefficients [19].

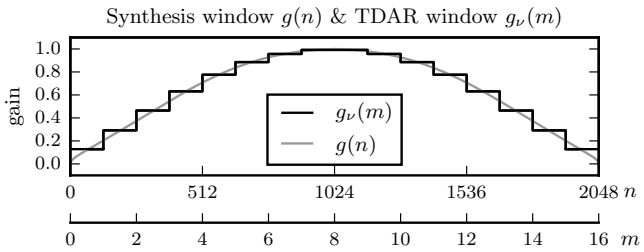

Fig. 2. Time Domain Aliasing Reduction (TDAR) coefficients approximation. Time Domain Aliasing Cancellation (TDAC) uses $2M$ synthesis coefficients $g(n)$. One solution for the $2N_\nu < 2M$ Time Domain Aliasing Reduction (TDAR) coefficients $g_\nu(m)$ can be approximated from them.

III. Time Domain Aliasing Reduction

Earlier we were able to show that a perfectly reconstructing subband merging filterbank using an MDCT for both analysis and subband merging transform is possible. Such a filterbank results in improved temporal compactness over rectangular frequency domain windowing. However, aliasing from (2) limits its benefits. To further improve temporal compactness, an analogue to the TDAC operation after the IMDCT (but missing after direct MDCT) needs to be found. In order to retain perfect reconstruction, this operation needs to be invertible.

For our MDCT transform kernel (3) TDAC can be rearranged to

$$\begin{aligned} x_i(n) &= \hat{x}_i(n)g(M/2+n) \\ &\quad - \hat{x}_{i-1}(M-1-n)g(3M/2+n), \end{aligned} \tag{12}$$

$$\begin{aligned} x_{i-1}(M-1-n) &= \hat{x}_i(n)g(M/2-1-n) \\ &\quad + \hat{x}_{i-1}(M-1-n)g(3M/2-1-n). \end{aligned} \tag{13}$$

by incorporating the IMDCT post-permutation into overlap add. Using this notation is required since the direct MDCT does not have a post-permutation stage in (subband) time domain and we only get M samples from the transform, while the operation used earlier (10) was performed after the IMDCT, which has a post-permutation stage and returns $2M$ samples. Both operations are identical, but in this form, TDAC is performed in-place. In-place TDAC can be written as

$$\begin{bmatrix} x_i(n) \\ x_{i-1}(M-1-n) \end{bmatrix} = \mathbf{A}_g(n) \begin{bmatrix} \hat{x}_i(n) \\ \hat{x}_{i-1}(M-1-n) \end{bmatrix} \tag{14}$$

with

$$\mathbf{A}_g(n) = \begin{bmatrix} g(M/2+n) & -g(3M/2+n) \\ g(M/2-1-n) & g(3M/2-1-n) \end{bmatrix} \tag{15}$$

for $0 \le n < M/2$. Note the use of $g(n)$, which has $2M$ window coefficients.

To reduce aliasing in the subband signal, this operation has to be performed on $\hat{y}_{\nu,i}(m)$. However, since we do not have M samples $\hat{x}_i(n)$ and $2M$ window coefficients $g(n)$ but only N_ν subband samples $\hat{y}_{\nu,i}(m)$ we need to find $2N_\nu$ synthesis window coefficients $g_\nu(m)$. Also, for TDAR to be invertible the resulting matrix $\mathbf{A}_{g_\nu}(m)$ must be nonsingular. Our experiments have shown that two very simple coefficient calculation schemes achieve good aliasing reduction with highly improved temporal compactness. Both methods assume the subband samples are equidistantly spread over the original frame and both produce $g_\nu(m)$ of length $2N_\nu$ from the available synthesis window $g(n)$:

1) For parametric windows such as *Sine* or *Kaiser Bessel Derived*, a shorter window of the same type can be defined.
2) For both parametric and tabulated windows, the window may be simply cut into $2N_\nu$ sections of equal size, allowing coefficients to be obtained using the root mean square value of each section:

$$g_\nu(m) = \sqrt{\frac{N_\nu}{M} \sum_{n=0}^{M/N_\nu-1} g^2\left(n + m\frac{M}{N_\nu}\right)} \quad 0 \le m < 2N_\nu. \tag{16}$$

This yields $2N_\nu$ window coefficients which we can use to perform TDAR by substituting $g_\nu(m)$ for $g(n)$ in (15) yielding

$$\begin{bmatrix} y_{\nu,i}(m) \\ y_{\nu,i-1}(N_\nu-1-m) \end{bmatrix} = \mathbf{A}_{g_\nu}(m) \begin{bmatrix} \hat{y}_{\nu,i}(m) \\ \hat{y}_{\nu,i-1}(N_\nu-1-m) \end{bmatrix} \tag{17}$$

with

$$\mathbf{A}_{g_\nu}(m) = \begin{bmatrix} g_\nu(N_\nu/2+m) & -g_\nu(3N_\nu/2+m) \\ g_\nu(N_\nu/2-1-m) & g_\nu(3N_\nu/2-1-m) \end{bmatrix}. \tag{18}$$

It can be shown that, if $g(n) = h(n)$ and $v(k) = w(k)$, i.e. both MDCTs are orthogonal, matrix $\mathbf{A}_{g_\nu}(m)$ is orthogonal and the overall pipeline is an orthogonal transform.

To calculate the inverse transform, first inverse TDAR is performed before applying the aforementioned synthesis filterbank

$$\begin{bmatrix} \hat{y}_{\nu,i}(m) \\ \hat{y}_{\nu,i-1}(N_\nu-1-m) \end{bmatrix} = \mathbf{A}_{g_\nu}^{-1}(m) \begin{bmatrix} y_{\nu,i}(m) \\ y_{\nu,i-1}(N_\nu-1-m) \end{bmatrix}. \tag{19}$$

Note that performing TDAR is optional. If a less temporally compact impulse response is acceptable, e.g. when coding a stationary signal frame, TDAR can be skipped. As TDAR can only be performed between two frames with identical subband layouts, switching from one subband layout to another requires TDAR to be skipped during transition.

A. TDAR Coefficient Optimization

The proposed TDAR coefficient derivation schemes are based on deriving $g_\nu(m)$ from the original synthesis window $g(n)$. However, provided the nonsingularity condition is not violated, any TDAR matrix $\mathbf{A}(m)$ is possible. This allows for numerical optimization.

As TDAR mainly works to improve temporal compactness (see Figures 3 and 4), the effective length of the impulse response l_{eff}^2 [20] should be used as the cost function and be minimized for each individual index m. In order not to violate the nonsingularity condition, each quadruple of coefficients in (18) must be jointly optimized.

This optimization is optional and for the following evaluation the coefficients were simply obtained using estimation method 1 from Section III.

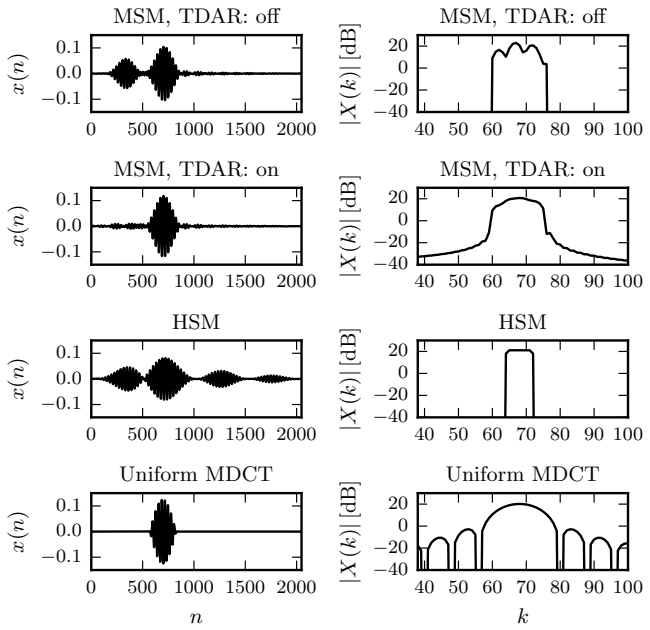

Fig. 3. Impulse and frequency response of a merged subband filter for several methods. Shown are the proposed MDCT Subband Merging (MSM) without Time Domain Aliasing Reduction (TDAR), MSM with TDAR, and Hadamard Subband Merging (HSM) as proposed in [4]. All methods are based on an original window length $2M = 2048$ and a merge factor $N = 8$. Also shown are the impulse and frequency response of the corresponding uniform MDCT with window length $2M/N = 256$ samples.

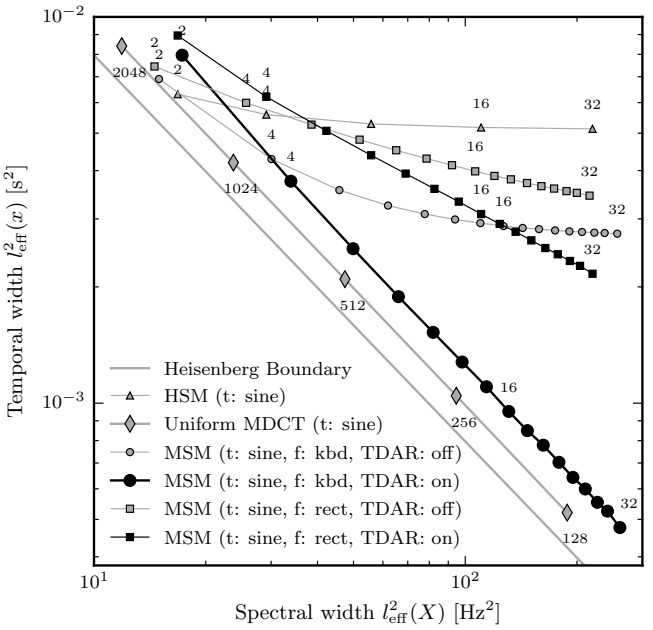

Fig. 4. Comparison of spectral and temporal compactness of several methods as presented in [3]. Shown are the proposed MDCT Subband Merging (MSM) and Hadamard Subband Merging (HSM) as proposed in [4]. All methods are based on an original MDCT with window length $2M = 2048$. Also shown are a uniform MDCT of variable window lengths and the theoretical lower bound as defined by the Heisenberg uncertainty relation. "t: sine" denotes *sine* temporal window, "f: kbd/rect" denotes *Kaiser Bessel derived* or *rectangular* spectral window, "TDAR: on/off" denotes enabled/disabled Time Domain Aliasing Reduction (TDAR). Inline labels denote window lengths for MDCT and merge factors for all others.

IV. Results

As introduced in Section I, the quantization noise is shaped by the impulse response shape. Consequently, we are mainly interested in temporally and spectrally compact impulse responses and will look at these attributes for our evaluation.

Fig. 3 clearly shows the poor temporal compactness of HSM. Due to rectangular windowing with no overlap, the very high spectral compactness can be seen. MSM, on the other hand, allows for a more temporally compact impulse response in the subband. However, aliasing is clearly visible as an unwanted impulse, left of the main inpulse, mirrored at the boundary at $n = 512$.

TDAR seems to significantly reduce most of these aliasing artifacts in the MSM subband. It can also be seen that spectral compactness of MSM with and without TDAR—both with overlap in frequency—are similar, with a gentler roll-off below -20 dB in case of TDAR. Fig. 4 shows that HSM offers severely limited time-frequency trade-off capabilities. For growing merge sizes, additional temporal resolution comes at an unproportionally high cost in spectral compactness, compromising the overall efficiency when used in an audio codec. This is the reason why the maximum merge factor in current codec implementations is limited to 8 [5], after which there is no more temporal resolution to be gained and the slope in the graph approaches zero.

MSM without TDAR ("TDAR: off") or with no spectral overlap ("f: rect") shows a slightly improved trade-off capabilities compared to HSM.

MSM with overlap and TDAR ("f: kbd, TDAR: on"), however, has a constant trade-off between temporal and spectral compactness. The slope never becomes zero, stays nearly constant and equal to the slope of an unmerged MDCT transform. This means all merge factors share the same high overall compactness and there is no merge factor at which further merging would become useless.

However, using TDAR for a merging factor $N_\nu = 2$ appears to decrease temporal compactness when compared to HSM or MSM without TDAR. TDAR can be disabled or the start/stop window method explained in Section II-C can be used to use HSM for bands with a merge factor $N_\nu = 2$, while using MSM for all higher merge factors.

For this analysis, a *sine* analysis window ("t: sine") and a *Kaiser Bessel derived* subband merging window ("f: kbd") showed the most compact results and thus were chosen.

V. Conclusion

We present a novel method for merging subbands of MDCT spectra and, using TDAR, subsequently reducing aliasing in each subband. The result is a non-uniform, critically sampled and perfect reconstructing filterbank with overlapping subband merging and freely selectable bandwidths. The introduced method provides much better temporal and spectral localization than state-of-the-art methods, even for larger merge factors, and therefore allows for a greater degree of freedom in the choice of bandwidths and consequently greater coding efficiency.

References

[1] Fernando C. Pereira and Touradj Ebrahimi, *The MPEG-4 Book*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2002.

[2] B. C. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.*, vol. 74, no. 3, pp. 750–753, Sep 1983.

[3] F. Bimbot, E. Camberlein, and P. Philippe, "Adaptive filter banks using fixed size MDCT and subband merging for audio coding - comparison with the MPEG AAC filter banks," in *Audio Engineering Society Convention 121*, Oct 2006.

[4] O.A. Niamut and R. Heusdens, "Subband merging in cosine-modulated filter banks," *Signal Processing Letters, IEEE*, vol. 10, no. 4, pp. 111–114, April 2003.

[5] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-quality, low-delay music coding in the Opus codec," in *Audio Engineering Society Convention 135*, Oct 2013.

[6] J. Kovacevic and M. Vetterli, "Perfect reconstruction filter banks with rational sampling factors," *Signal Processing, IEEE Transactions on*, vol. 41, no. 6, pp. 2047–2066, June 1993.

[7] B.-J. Yoon, H. Malvar, and R. Malvar, "The design of nonuniform lapped transforms," Tech. Rep., September 2005.

[8] T. Necciari, P. Balazs, N. Holighaus, and P.L. Sondergaard, "The ERBlet transform: An auditory-based time-frequency representation with perfect reconstruction," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 498–502.

[9] O. Derrien, T. Necciari, and P. Balazs, "A quasi-orthogonal, invertible, and perceptually relevant time-frequency transform for audio coding," in *EUSIPCO*, Nice, France, Aug. 2015.

[10] H. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts," *Signal Processing, IEEE Transactions on*, vol. 46, no. 4, pp. 1043–1053, Apr 1998.

[11] G. D. T. Schuller and M. J. T. Smith, "New framework for modulated perfect reconstruction filter banks," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1941–1954, Aug 1996.

[12] ISO, "Enhanced low delay AAC," ISO 14496–3:2005/Amd 9:2008, International Organization for Standardization, Geneva, Switzerland, 2008.

[13] J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1153–1161, Oct 1986.

[14] J. Princen, A. Johnson, and A. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87.*, Apr 1987, vol. 12, pp. 2161–2164.

[15] R. K. Chivukula, Y. A. Reznik, and V. Devarajan, "Efficient algorithms for MPEG-4 AAC-ELD, AAC-LD and AAC-LC filterbanks," in *2008 International Conference on Audio, Language and Image Processing*, July 2008, pp. 1629–1634.

[16] H. Malvar, "Fast algorithm for modulated lapped transform," *Electronics letters*, vol. 27, no. 9, pp. 775–776, 1991.

[17] V. Britanak and H. J. Lincklaen Arriëns, "Fast computational structures for an efficient implementation of the complete {TDAC} analysis/synthesis MDCT/MDST filter banks," *Signal Processing*, vol. 89, no. 7, pp. 1379 – 1394, 2009.

[18] B. Edler, "Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen," *Frequenz*, vol. 43, pp. 252–256, Sept. 1989.

[19] J. Lecomte, P. Gournay, R. Geiger, B. Bessette, and M. Neuendorf, "Efficient cross-fade windows for transitions between LPC-based and non-LPC based audio coding," in *Audio Engineering Society Convention 126*, May 2009.

[20] A. Papoulis, *Signal analysis*, Electrical and electronic engineering series. McGraw-Hill, New York, San Francisco, Paris, 1977.