

ON THE ENERGY COMPACTION PERFORMANCE OF THE BIORTHOGONAL MODULATED LAPPED TRANSFORM

Jani Tuisku

Signal Processing Laboratory, Tampere University of Technology
P.O.Box 553, FIN-33101 Tampere, Finland

ABSTRACT

The use of orthogonal lapped transforms, and in particular the modulated lapped transform (MLT), is common practice in audio coding. A more general class of transforms, namely the biorthogonal lapped transforms, have not been popularized yet. In this paper the particular case of biorthogonal MLT is studied more closely. The biorthogonal MLT is based on selection of the transform window for either the analysis or the synthesis transform, and computing the other window to satisfy the biorthogonality requirement. Different window functions, some of which yield potential biorthogonal lapped transforms for audio coding, are presented and compared from the energy compaction point of view.

1. INTRODUCTION

The modulated lapped transform (MLT) [1] is one of the most common, if not the most common, transform methods used in the field of audio coding. It finds its application in nearly all current audio coding standards. The modified discrete cosine transform (MDCT) is a commonly used alias for the MLT, while the time-domain aliasing cancellation (TDAC) filterbank with odd channel stacking is practically identical, but somewhat older concept.

The MLT is an orthogonal lapped transform, with M subbands and basis functions of length $L = 2M$. The name of the MLT is easily understood, considering two facts about its design. Firstly, it is based on a lowpass prototype filter, or the window function $h[n]$, that is cosine modulated to different center frequencies to yield a set of basis functions of length L . Secondly, the MLT processes its input with an overlap of M samples between adjacent transform blocks.

In matrix notation, the MLT can be described by using matrix multiplications by an orthogonal transform matrix \mathbf{P} , of size $L \times M$, with the MLT basis functions as its columns. In performing the MLT, each input block of L time-domain samples \mathbf{x} is mapped to

a set of M transform coefficients \mathbf{X} , by multiplying the block with \mathbf{P}^T :

$$\mathbf{X} = \mathbf{P}^T \mathbf{x}.$$

Time-domain samples can then be reconstructed from the transform coefficients, by performing the inverse MLT, which is a multiplication by \mathbf{P} :

$$\mathbf{x} = \mathbf{P} \mathbf{X}.$$

Because the transform blocks are lapped, the reconstruction actually becomes the infinite sum of inverse MLTs over all transform blocks. In order to guarantee perfect reconstruction, \mathbf{P} must satisfy the orthogonality condition

$$\mathbf{P}^T \mathbf{W}^m \mathbf{P} = \delta(m) \mathbf{I}, \quad m = 0, 1, \quad (1)$$

where \mathbf{W} is the one block forward shift matrix.

Basis functions of the MLT, that potentiate perfect reconstruction, are defined as

$$p_{n,k} = h[n] \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right],$$

for $k = 0, 1, \dots, M-1$, and $n = 0, 1, \dots, L-1$. The factor $\sqrt{2/M}$ is used for normalization, and $h[n]$ is the window function.

The window function is generally assumed to be evenly symmetric:

$$h[L-1-n] = h[n]. \quad (2)$$

In order for the transform to be perfectly reconstructable, $h[n]$ then has to satisfy the requirement:

$$h^2[n] + h^2[n+M] = 1. \quad (3)$$

A common choice for the window function is the half-sine window

$$\sin[n] = \pm \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{L} \right], \quad (4)$$

where the sign can be chosen according to preference on the sign of the lowpass subband. This window, guaranteeing perfect reconstruction, is used for example in Layer 3 of MPEG-1 and MPEG-2 standards [2], [3].

2. THE BIORTHOGONAL MODULATED LAPPED TRANSFORM

In the case of MLT the transform matrix \mathbf{P} is an orthogonal matrix. This means that the forward and inverse transform matrices are restricted to be transposes of each other. By convention \mathbf{P}^T is used for the forward transform and \mathbf{P} for the inverse transform.

More freedom in the design of the transform basis functions can, however, be introduced by using two different matrices as the transform matrices, one for the forward and one for the inverse transform matrix. Let us denote the forward transform matrix by \mathbf{H}^T and the inverse transform matrix by \mathbf{G} .

Corresponding to the condition (1), the transform matrices must satisfy the biorthogonality condition

$$\mathbf{H}^T \mathbf{W}^m \mathbf{G} = \delta(m) \mathbf{I}, \quad m = 0, 1.$$

We can now otherwise keep to the definition of the MLT basis functions $p_{n,k}$ above, but make use of two different window functions. Let $h_0[n]$ be the analysis window and $h_1[n]$ the synthesis window. Under the requirement of symmetry (2), the window functions now have to satisfy [5]:

$$h_0[n]h_1[n] + h_0[n+M]h_1[n+M] = 1$$

and

$$h_0[n]h_1[n+M] - h_0[n+M]h_1[n] = 0.$$

One of the windows can be fixed and the other solved according to a choice. Related to this work it is worth more interest to be able to choose the analysis window. The synthesis window then becomes

$$h_1[n] = \frac{h_0[n]}{h_0^2[n] + h_0^2[n+M]}. \quad (5)$$

The analysis window can now be tuned as desired, for example to maximize the energy compaction of the forward transform. The synthesis window can then be computed according to (5), still being able to maintain perfect reconstruction.

3. SOME WINDOW FUNCTIONS

In this chapter some window functions of interest are presented. Unless otherwise noted, all the windows are expressed in such a form that results in a symmetric window on the range $n = 0, 1, \dots, L-1$.

3.1. $\sin[n]$ Window

The half-sine window (4) is hereafter referred to as the $\sin[n]$ window. It serves as a good reference point as it is used in the popular MPEG Layer 3.

3.2. $\sin^\alpha[n]$ Window

In order to see the difference between orthogonal and biorthogonal MLT, we can consider the generalized power of the $\sin[n]$ window, $\sin^\alpha[n]$.

3.3. Kaiser-Bessel Window

Although the Kaiser-Bessel (KB) window is a well known window function, its use is not always totally unambiguous. Various adaptations of the definition found in [6] exist. In this work the Kaiser-Bessel window is defined as

$$w[n, \alpha] = \frac{I_0 \left(\pi \alpha \sqrt{1 - \left(\frac{2n+1}{L} - 1 \right)^2} \right)}{I_0(\pi \alpha)}, \quad (6)$$

where $I_0()$ is the zeroth order modified Bessel function of the first kind. The window shape can be modified by varying α , which determines the time-bandwidth product of the window.

3.4. Kaiser-Bessel Derived Window

Disregarding the effect of transform block switching, MPEG-2 AAC [4] makes use of the so called Kaiser-Bessel derived (KBD) window

$$W[n, \alpha] = \frac{\sum_{p=0}^{L-1} s[p]w[n-p, \alpha]}{\sqrt{\sum_{p=0}^{L/2} w[p, \alpha]}}, \quad (7)$$

with

$$s[p] = \begin{cases} 1, & 0 \leq p < M \\ 0, & \text{otherwise} \end{cases},$$

where $w[n, \alpha]$ is the Kaiser-Bessel window function.

3.5. Modified Kaiser-Bessel Window

The Kaiser-Bessel window does not decay to zero at its boundaries. This has been observed in [7], with a suggestion to rectify the problem in the form

$$\hat{w}[n, \alpha] = \frac{I_0 \left(\pi \alpha \sqrt{1 - \left(\frac{2n+1}{L} - 1 \right)^2} \right) - 1}{I_0(\pi \alpha) - 1}. \quad (8)$$

This modified Kaiser-Bessel (MBK) window does decay to zero at the boundaries and has better asymptotic behavior than the original Kaiser-Bessel window.

3.6. Modified Kaiser-Bessel Derived Window

It is now worthwhile to consider a further modification based on the previous windows. This would be logically named the modified Kaiser-Bessel derived (MKBD) window, and results from applying the modified Kaiser-Bessel window formula in the Kaiser-Bessel derived window equation. In addition a general exponent β can be used, in the spirit of the $\sin^\alpha[n]$ window, instead of the square root found in the KBD formula:

$$\hat{W}[n, \alpha] = \left[\frac{\sum_{p=0}^{L-1} s[p] \hat{w}[n-p, \alpha]}{\sum_{p=0}^{L/2} \hat{w}[p, \alpha]} \right]^\beta. \quad (9)$$

The decay properties of the MKBD window should be better on the expectation than those of the original KBD window, as a consequence of using a prototype window with improved decay.

3.7. MLBT Window of Malvar

Previously Malvar [8] has proposed the use of a window of the form

$$h_s[n] = \frac{1 - \cos \left[\left(\frac{n+1}{M} \right)^\alpha \pi \right] + \beta}{2 + \beta}, \quad (10)$$

with $n = 0, 1, \dots, M-1$, for the synthesis window of the biorthogonal MLT. It may be preferred, however, to apply the window for the design of analysis filters instead, as it appears to result almost systematically in higher coding gain. This alternative is considered here.

4. CODING GAIN

Coding gain is a commonly used theoretical figure of merit for the energy compaction performance of transforms, and is a very useful tool in comparing different transforms. It expresses the ratio of the quantization error variance of pulse code modulated (PCM) quantization to that of transform coded quantization, under the assumption of using scalar quantizers at high bit rate. Higher coding gain values are indicative of better transform energy compaction performance.

The well known coding gain formula of [9] only applies for orthogonal transforms, which have the property of preserving the signal variance. Biorthogonal transforms, however, do not preserve the signal variance. In [10], the coding gain formula has been generalized using constants A_k and B_k to account for the

change in signal variance by the forward and inverse transform basis functions, respectively. A_k can be defined as

$$A_k = \sigma_{X_k}^2 / \sigma_x^2,$$

and assuming quantization noise from different transform channels to be uncorrelated and white,

$$B_k = \frac{1}{M} \sum_{n=0}^{L-1} g_{n,k}^2.$$

Above, $\sigma_{X_k}^2$ is the variance of the transform coefficient X_k , while σ_x^2 is the variance of the input signal x . Finally the inverse transform basis functions are defined by $g_{n,k}$.

Generalized coding gain, that applies for biorthogonal as well as orthogonal transforms, now has the expression

$$G_{TC} = \frac{1}{M} \left(\prod_{k=0}^{M-1} A_k B_k \right)^{-1/M}.$$

5. RESULTS

Biorthogonal modulated lapped transforms stemming from the presented window functions were tested using the popular ‘‘Harpichord’’ sample. It provides a demanding test signal with wide spectral content and sharp transients. Generalized coding gain was computed for different transform lengths and over a selected feasible range of window parameters. The results indicating maximum coding gain in dB for a given window and transform length are summarized in Table 1.

The $\sin[n]$ window performs clearly very poorly for short transform lengths, improving at around 256 samples. The $\sin^\alpha[n]$ window is a clear improvement, resulting in an overall improvement and the best performance for long transform blocks. Smaller exponent values and consequently wider windows would appear to be more optimum for longer transform lengths.

The KB window is totally out of the competition, but the MKB window provides a generous improvement. The MKBD seems to be clearly better than the MKB window, resulting in the best performance over intermediate transform lengths. Finally the window (10) provides a good overall performance.

The results show that the $\sin[n]$ window should not be used for short transform lengths and that the KBD window can be outperformed using simple modifications such as the MKB or the MKBD window. Although the overall differences between the best windows are not very big, the presented windows may prove more valuable in other applications.

Window	Transform length			
	32	64	128	256
$\sin[n]$	10.55	12.07	13.16	14.35
$\sin^\alpha[n]$	1.2	1.9	1.8	1.5
	10.57	12.57	13.89	14.67
KB (6)	1.8	2.2	2.4	2.3
	10.43	12.12	13.17	13.80
KBD (7)	2.0	3.4	3.8	3.7
	10.47	12.64	14.00	14.74
MKB (8)	1.4	1.7	1.8	1.4
	10.65	12.53	13.69	14.54
MKBD (9)	0.8	2.7	2.8	2.4
	0.0	0.6	0.6	0.6
	10.57	12.76	14.03	14.75
Malvar (10)	0.8	1.0	0.9	0.8
	0.0	0.0	0.0	0.0
	10.57	12.62	13.92	14.72

Window	Transform length			
	512	1024	2048	4096
$\sin[n]$	15.28	15.99	16.99	18.04
$\sin^\alpha[n]$	1.2	1.0	0.8	0.9
	15.32	15.99	17.03	18.05
KB (6)	2.2	2.1	2.0	1.9
	14.42	15.11	15.95	17.34
KBD (7)	3.4	2.6	1.9	1.5
	15.28	15.90	16.89	18.00
MKB (8)	0.9	0.3	0.1	0.6
	15.29	16.00	17.02	18.04
MKBD (9)	1.0	0.1	0.7	0.7
	0.6	0.5	0.4	0.5
	15.31	15.99	17.01	18.04
Malvar (10)	0.7	0.6	0.5	0.6
	0.0	0.0	0.0	0.0
	15.34	16.00	17.02	18.01

Table 1: Results of simulations. Each table cell consists of the window parameters (α, β), when applicable, and the maximum coding gain (dB) found. The boxed values indicate the highest coding gain for a given transform length.

6. CONCLUSION

The concept of biorthogonal modulated lapped transform has been presented together with some potential window functions. To the author's knowledge some of the windows considered have not been previously studied in relation to audio coding. This is probably due to the fact that biorthogonal techniques in general have not been researched very deeply inside the field.

The experiments carried out promise somewhat

improvement in coding gain over the existing transform methods that are currently employed in audio coding standards. Some of the window functions studied, in particular the proposed MKBD window, seem to be interesting and may appear valuable in other applications such as frequency analysis.

7. REFERENCES

- [1] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, MA: Artech House, 1992.
- [2] ISO/IEC JTC1/SC29/WG11, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbit/s, International Standard ISO/IEC 11172-3, Part 3: Audio," 1993.
- [3] ISO/IEC JTC1/SC29/WG11, "Information Technology: Generic Coding of Moving Pictures and Associated Audio: Audio, International Standard ISO/IEC 13818-3," 1994.
- [4] ISO/IEC JTC1/SC29/WG11, "MPEG-2 DIS 13818-7 (MPEG Advanced Audio Coding, AAC)," 1996.
- [5] S. Cheung and J. S. Lim, "Incorporation of Bi-orthogonality into Lapped Transforms for Audio Compression," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, May 1995, pp. 3079–3082.
- [6] F. J. Harris, "On the Use of Windows for Harmonic Analysis of the Discrete Fourier Transform," *Proceedings of the IEEE*, vol. 66, pp. 51–83, January 1975.
- [7] N. C. Geckinli and D. Yavuz, "Some Novel Windows and a Concise Tutorial Comparison of Window Families," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 6, pp. 501–507, December 1978.
- [8] H. S. Malvar, "Biorthogonal and Nonuniform Lapped Transforms for Transform Coding with Reduced Blocking and Ringing Artifacts," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1043–1053, April 1998.
- [9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice Hall, 1984.
- [10] S. O. Aase, *Image Subband Coding Artifacts: Analysis and Remedies*, PhD thesis, The Norwegian Institute of Technology, 1993.