CONSIDERATIONS ON WINDOW SWITCHING WITH THE BIORTHOGONAL MODULATED LAPPED TRANSFORM

Jani Tuisku

Signal Processing Laboratory, Tampere University of Technology
P.O.Box 553, FIN-33101 Tampere, Finland
jani.tuisku@cs.tut.fi

ABSTRACT

The modulated lapped transform (MLT) can be regarded as the fundamental transform in the field of audio coding. The MLT is, however, an orthogonal tranform, which poses the limitation of the synthesis transform matrix being simply the transpose of the analysis transform matrix. This requires the analysis and synthesis window functions to be the same. It is possible to generalize the MLT into the biorthogonal modulated lapped transform (BMLT), and gain more flexibility in the choice of the analysis and synthesis window functions. It is further possible to adapt the BMLT window functions for example to tune the energy compaction of the analysis transform. This paper addresses the requirements of perfect reconstruction when the BMLT transform windows are varied in time.

1. INTRODUCTION

Audio coding is a process that is applied for sampled time-domain audio waveforms. The process consists of the application of a pair of an encoding operation and a corresponding decoding operation. In the encoding operation the time-domain audio waveform is turned into a coded representation that is more efficient in terms of the number of bits required to represent a segment of the waveform. The inverse operation of audio decoding can be used to recover the time-domain representation of the audio signal from the coded one. Depending on the lossyness of the encoding and decoding operations, the reconstructed time-domain waveform can either be identical to the original one or not. Lossless coding does not introduce any coding errors, which corresponds to the reconstructed waveform being identical to the original one. This imposes a limitation on the degree of bit rate reduction achievable. Lossy coding is more popular than lossless coding, because it achieves higher compression by allowing errors in the reconstructed waveform.

Perceptual audio coding is an important subclass of lossy audio coding, that makes use of the deficiencies of human auditory perception. A perceptual audio encoder can be considered to consist of four basic building blocks. These include the analysis filterbank, quantization and coding, psychoacoustical model, and encoding of the bitstream. The task of the analysis filterbank is to decompose the time-domain signal into short blocks of subsampled spectral components for further processing. These spectral components are quantized and coded according to the resolution required by human auditory perception. This resolution information is provided by a psychoacoustical model that seeks to model the psychoacoustical masking behavior of the signal to be coded. The final stage in the encoding process is the encoding of the coded spectral components and necessary side information into a bitstream with a lower bit rate than is used for the time-domain signal.

The analysis and synthesis filterbanks in perceptual audio coding methods are generally implemented using so called lapped transforms. Time-domain samples of audio signals are processed in short blocks which are windowed using bell-shaped transform windows that decay towards zero at the block boundaries. Adjacent transform blocks overlap each other, hence the name lapped transforms. The advantage of the overlapping of the transform blocks is that blocking effects common to lossy coding methods are avoided. The use of lapped transforms can therefore be considered as the fundamental element of transform based audio coding. Probably the most fundamental lapped transform, with respect to audio coding, is the modulated lapped transform (MLT) [1], which is also known under the name modified discrete cosine transform (MDCT). The MLT is an orthogonal lapped transform, the application of which results in M transform coefficients for every $L = 2M$ input signal samples. It finds its use for example in the MPEG audio coding standards [2], [3], [4], [5].

It is possible to change the definition of the MLT somewhat and end up in a more general biorthogonal lapped transform [6]. This biorthogonal modulated lapped transform (BMLT) has more degrees of freedom in its design compared to the MLT. In particular the shape of the transform window can be adapted, which can be used for optimizing the performance of the transform. The energy compaction performance of the analysis filterbank is known to be an important parameter in coding applications, because it can be linked to the performance of subsequent quantization and coding stages. The effect of transform window adaptation on the energy compaction performance of the biorthogonal MLT has recently been considered in [7] and [8]. The results show that optimization of the analysis window can be used to improve the transform performance.

Knowledge of psychoacoustics has been widely utilized in the field of perceptual audio coding. Psychoacoustical models have been designed, and are commonly employed, to characterize audio signals in terms of their frequency domain masking behavior. This kind of an approach is well justified in stationary modeling of audio signals in terms of their auditory masking properties. Its application is, however, more than arguable for transient parts of audio signals where abrupt changes in signal component intensities occur. Generally applicable masking models for transients have not been introduced because the time domain masking phenomenon is very complex to be modeled.

In transform based audio coding, transients are commonly treated by switching the encoder to use a shorter transform length over the transient parts, than the stationary parts of signals. This procedure of window switching is applied in order to limit so called pre-echos, which are particularly annoying audible artifacts caused by quantization noise exceeding the auditory masking threshold. The transform window of a lapped transform has to satisfy certain conditions in order for the transform to be perfectly reconstructible. In the situation of window switching, these conditions are known for orthogonal lapped transforms [9], [10], and are also in application in all the above mentioned audio coding standards. For biorthogonal lapped transforms the conditions of perfect reconstruction, in the situation of window switching, are yet unpublished. Such conditions are important, because they are related to the transient performance of biorthogonal lapped transforms. In this paper the conditions are derived, making use of the results in [9] and [11].

2. THE MODULATED LAPPED TRANSFORM

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 reconstruced 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 becomes an infinite sum of inverse MLTs over all transform blocks. 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 $k = 0, 1, \ldots, M-1$, and $n = 0, 1, \ldots, L-1$. 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,

$$h(L-1-n) = h(n), \tag{1}$$

$n = 0, 1, \ldots, M-1$.

The $p_{n,k}$ above define the basis matrix $\mathbf{P}$ of the MLT. Here $\mathbf{P}$ is considered to consist of two $M \times M$ submatrices,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}'_0 \\ \mathbf{P}'_1 \end{bmatrix},$$

with $\mathbf{P}'_0 = \mathbf{P}_0\mathbf{W}_0$ and $\mathbf{P}'_1 = \mathbf{P}_1\mathbf{W}_1$. This notation has the advantage of separating the windowing operation from the rest of the transform matrix. Here $\begin{bmatrix} \mathbf{P}_0^T & \mathbf{P}_1^T \end{bmatrix}^T$ is the basis matrix disregarding the window function, that is $p_{n,k}$ with $h(n) = 1$, $n = 0, 1, \ldots, L-1$. The window function is included in the windowing matrices $\mathbf{W}_0$ and $\mathbf{W}_1$ which are defined as

$$\begin{aligned} \mathbf{W}_0 &= \begin{bmatrix} h(0) & & & \mathbf{0} \\ & h(1) & & \\ & & \ddots & \\ \mathbf{0} & & & h(N-1) \end{bmatrix} \\ \mathbf{W}_1 &= \begin{bmatrix} h(N) & & & \mathbf{0} \\ & \ddots & & \\ & & h(L-2) & \\ \mathbf{0} & & & h(L-1) \end{bmatrix}. \end{aligned}$$

The submatrices $\mathbf{P}_0$ and $\mathbf{P}_1$ have a couple of useful properties which can be expressed as

$$\mathbf{P}_0^T\mathbf{P}_1 = \mathbf{0} \tag{2}$$

$$\mathbf{P}_1^T\mathbf{P}_0 = \mathbf{0} \tag{3}$$

$$\mathbf{P}_0^T\mathbf{P}_0 = \mathbf{I} - \mathbf{J} \tag{4}$$

$$\mathbf{P}_1^T\mathbf{P}_1 = \mathbf{I} + \mathbf{J}. \tag{5}$$

3. THE BIORTHOGONAL MODULATED LAPPED TRANSFORM

The MLT transform matrix $\mathbf{P}$ is an orthogonal matrix. This means that the analysis and synthesis transform matrices are restricted to be transposes of each other. By convention $\mathbf{P}^T$ is used for the analysis transform and $\mathbf{P}$ for the synthesis 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 analysis and one for the synthesis.

Let us define the analysis transform matrix as

$$\mathbf{H}^T = \begin{bmatrix} \mathbf{H}_0^T & \mathbf{H}_1^T \end{bmatrix},$$

with $\mathbf{H}_0 = \mathbf{P}_0\mathbf{W}_{H_0}$ and $\mathbf{H}_1 = \mathbf{P}_1\mathbf{W}_{H_1}$. Similarly define the synthesis transform matrix as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix},$$

with $\mathbf{G}_0 = \mathbf{P}_0\mathbf{W}_{G_0}$ and $\mathbf{G}_1 = \mathbf{P}_1\mathbf{W}_{G_1}$.

Now both the transform matrices make use of the same modulating cosines, $\mathbf{P}_0$ and $\mathbf{P}_1$, but different window functions. The analysis window is denoted as $\mathbf{W}_H$ and the synthesis window as $\mathbf{W}_G$. The result is called the biorthogonal modulated lapped transform.

4. PERFECT RECONSTRUCTION OF THE BMLT

Let us define our input signal as an infinite vector of samples,

$$\tilde{\mathbf{x}} = [\ldots, x(-2), x(-1), x(0), x(1), x(2), \ldots]^T.$$

The BMLT of this sample vector can now be expressed as

$$\tilde{\mathbf{X}} = \tilde{\mathbf{H}}^T\tilde{\mathbf{x}}, \tag{6}$$

where $\tilde{\mathbf{H}}^T$ is an infinite transform matrix. It is possible to recover the input sample vector $\tilde{\mathbf{x}}$ from the transform coefficients $\tilde{\mathbf{X}}$ by applying the inverse BMLT,

$$\tilde{\mathbf{x}} = \tilde{\mathbf{G}}\tilde{\mathbf{X}}, \tag{7}$$

where $\tilde{\mathbf{G}}$ is the infinite inverse transform matrix.

Assuming that the analysis transform matrix $\mathbf{H}^T$ and the synthesis transform matrix $\mathbf{G}$ are used for all the transform blocks, we can express the infinite transform matrices in terms of the previously defined submatrices as

$$\tilde{\mathbf{H}}^T = \begin{bmatrix} \ddots & & & \\ \mathbf{H}_0^T & \mathbf{H}_1^T & & \\ & \mathbf{H}_0^T & \mathbf{H}_1^T & \\ & & & \ddots \end{bmatrix}$$

and

$$\tilde{\mathbf{G}} = \begin{bmatrix} \ddots & & & \\ & \mathbf{G}_0 & & \\ & \mathbf{G}_1 & \mathbf{G}_0 & \\ & & \mathbf{G}_1 & \\ & & & \ddots \end{bmatrix}.$$

It is clear from equations (6) and (7) that perfect reconstruction of the BMLT of infinite length requires that

$$\tilde{\mathbf{G}}\tilde{\mathbf{H}}^T = \tilde{\mathbf{I}},$$

which is equivalent to

$$\begin{aligned} \mathbf{G}_0\mathbf{H}_0^T + \mathbf{G}_1\mathbf{H}_1^T &= \mathbf{I} \\ \mathbf{G}_0\mathbf{H}_1^T = \mathbf{G}_1\mathbf{H}_0^T &= \mathbf{0}, \end{aligned}$$

which in turn is equivalent to

$$\mathbf{G}_0^T\mathbf{H}_0 + \mathbf{G}_1^T\mathbf{H}_1 = \mathbf{I} \tag{8}$$

$$\mathbf{G}_0^T\mathbf{H}_1 = \mathbf{G}_1^T\mathbf{H}_0 = \mathbf{0}. \tag{9}$$

Next the requirements for perfect reconstruction are studied starting from (8) and (9). Equations (9) can be verified to be satisfied,

$$\begin{aligned} \mathbf{G}_0^T\mathbf{H}_1 &= (\mathbf{P}_0\mathbf{W}_{G_0})^T\mathbf{P}_1\mathbf{W}_{H_1} \\ &= \mathbf{W}_{G_0}^T\mathbf{P}_0^T\mathbf{P}_1\mathbf{W}_{H_1} = \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} \mathbf{G}_1^T\mathbf{H}_0 &= (\mathbf{P}_1\mathbf{W}_{G_1})^T\mathbf{P}_0\mathbf{W}_{H_0} \\ &= \mathbf{W}_{G_1}^T\mathbf{P}_1^T\mathbf{P}_0\mathbf{W}_{H_0} = \mathbf{0}, \end{aligned}$$

due to (2) and (3) correspondingly. Equation (8) gets the form

$$\begin{aligned} &\mathbf{G}_0^T\mathbf{H}_0 + \mathbf{G}_1^T\mathbf{H}_1 = \\ &\quad (\mathbf{P}_0\mathbf{W}_{G_0})^T\mathbf{P}_0\mathbf{W}_{H_0} + (\mathbf{P}_1\mathbf{W}_{G_1})^T\mathbf{P}_1\mathbf{W}_{H_1} = \\ &\quad \mathbf{W}_{G_0}^T\mathbf{P}_0^T\mathbf{P}_0\mathbf{W}_{H_0} + \mathbf{W}_{G_1}^T\mathbf{P}_1^T\mathbf{P}_1\mathbf{W}_{H_1} = \mathbf{I}, \end{aligned}$$

which simplifies due to properties (4) and (5) into the form

$$\mathbf{W}_{G_0}\mathbf{W}_{H_0} + \mathbf{W}_{G_1}\mathbf{W}_{H_1} = \mathbf{I}.$$

The result can be expressed in terms of the analysis and synthesis window functions as

$$h_G(i)h_H(i) + h_G(i+M)h_H(i+M) = 1, \tag{10}$$

$i = 0, 1, \ldots, M-1$. This is what has previously been published in [6]. It is, however, worthwile to note that the above analysis proves that the orthogonality condition (9) is satisfied independent of the analysis and synthesis windows chosen, which is different from what was found in [6]. This interesting result applies for the MLT as well as for the BMLT.

5. ASYMMETRICAL WINDOW SWITCHING

Let us now consider a case of switching both the analysis and synthesis transform windows at a certain point of time. First the analysis window H_A and the synthesis window G_A are used. At the middle of a certain transform block the windows are changed to H_B and G_B, correspondingly (see Figure 1 (a) for a simple illustration). The infinite analysis transform matrix now becomes

$$\tilde{\mathbf{H}}^T = \begin{bmatrix} \ddots & & & & & \\ \mathbf{H}_{A0}^T & \mathbf{H}_{A1}^T & & & & \\ & \mathbf{H}_{A0}^T & \mathbf{H}_{A1}^T & & & \\ & & \mathbf{H}_{A0}^T & \mathbf{H}_{B1}^T & & \\ & & & \mathbf{H}_{B0}^T & \mathbf{H}_{B1}^T & \\ & & & & \mathbf{H}_{B0}^T & \mathbf{H}_{B1}^T \\ & & & & & \ddots \end{bmatrix}$$

and the infinite synthesis transform matrix becomes

$$\tilde{\mathbf{G}} = \begin{bmatrix} \ddots & & & & & \\ \mathbf{G}_{A0} & & & & & \\ \mathbf{G}_{A1} & \mathbf{G}_{A0} & & & & \\ & \mathbf{G}_{A1} & \mathbf{G}_{A0} & & & \\ & & \mathbf{G}_{B1} & \mathbf{G}_{B0} & & \\ & & & \mathbf{G}_{B1} & \mathbf{G}_{B0} & \\ & & & & \mathbf{G}_{B1} & \\ & & & & & \ddots \end{bmatrix}.$$

It can be verified that the orthogonality conditions corresponding to (9) are satisfied due to (2) and (3).

Perfect reconstruction now requires the orthogonality condition (8) to be fulfilled by the submatrices corresponding to both the window pairs

$$\begin{aligned} \mathbf{G}_{A0}^T\mathbf{H}_{A0} + \mathbf{G}_{A1}^T\mathbf{H}_{A1} &= \mathbf{I} & (11) \\ \mathbf{G}_{B0}^T\mathbf{H}_{B0} + \mathbf{G}_{B1}^T\mathbf{H}_{B1} &= \mathbf{I}. & (12) \end{aligned}$$

With a similar analysis it can be verified that if the transform windows are switched in the middle of each transform block, perfect reconstruction is satisfied if the submatrices of all the overlapping transform blocks satisfy the condition (8). In terms of transform windows this corresponds to the requirement that the analysis and synthesis windows of overlapping transform blocks have to satisfy (10).

6. SYMMETRICAL WINDOW SWITCHING

It is also possible to consider switching the transform windows at the start of a certain transform block (see Figure 1 (b)). Using the previous notations of swithing the analysis

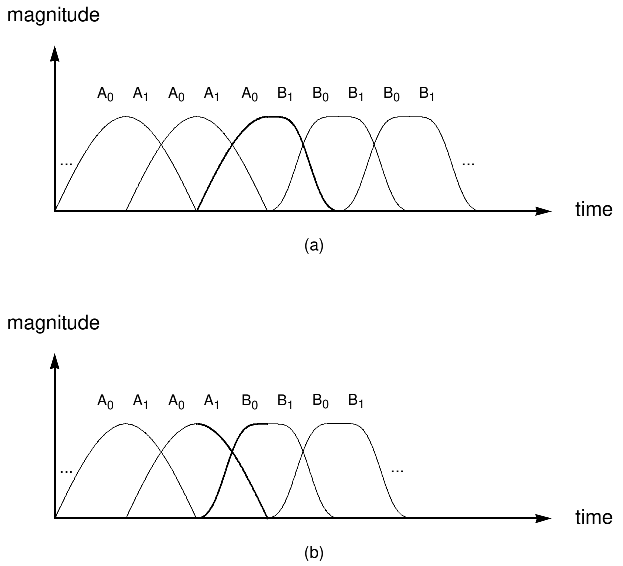

Figure 1: *Transform window switching. The window parts shown in thick lines indicate the window switching region. (a) Asymmetrical switching. (b) Symmetrical switching.*

window from H_A to H_B and the synthesis window from G_A to G_B, the infinite analysis transform matrix now becomes

$$\tilde{\mathbf{H}}^T = \begin{bmatrix} \ddots & & & & \\ \mathbf{H}_{A0}^T & \mathbf{H}_{A1}^T & & & \\ & \mathbf{H}_{A0}^T & \mathbf{H}_{A1}^T & & \\ & & \mathbf{H}_{B0}^T & \mathbf{H}_{B1}^T & \\ & & & \mathbf{H}_{B0}^T & \mathbf{H}_{B1}^T \\ & & & & \ddots \end{bmatrix}$$

while the infinite synthesis transform matrix becomes

$$\tilde{\mathbf{G}} = \begin{bmatrix} \ddots & & & & \\ \mathbf{G}_{A0} & & & & \\ \mathbf{G}_{A1} & \mathbf{G}_{A0} & & & \\ & \mathbf{G}_{A1} & \mathbf{G}_{B0} & & \\ & & \mathbf{G}_{B1} & \mathbf{G}_{B0} & \\ & & & \mathbf{G}_{B1} & \\ & & & & \ddots \end{bmatrix}.$$

It appears that in addition to the previous orthogonality conditions of submatrices corresponding to window pairs, (11) and (12), there is now an additional orthogonality condition related to the submatrices of overlapping transforms with different window functions,

$$\mathbf{G}_{B0}^T\mathbf{H}_{B0} + \mathbf{G}_{A1}^T\mathbf{H}_{A1} = \mathbf{I},$$

that needs to be satisfied for perfect reconstruction.

In terms of the analysis and synthesis window functions this can be expressed as

$$h_{G_B}(i)h_{H_B}(i) + h_{G_A}(i+M)h_{H_A}(i+M) = 1,$$

$i = 0, 1, \ldots, M-1$.

It can be verified that a corresponding requirement applies to all the analysis and synthesis windows of overlapping transform blocks with different windows.

7. DISCUSSION

Current audio coding standards only make use of the MLT and not the BMLT. The use of BMLT provides more degrees of freedom in the choice of the transform windows, which could be used to enhance the performance of audio coding methods. This increased variety in design choices provides more potential also in terms of window switching.

The MPEG audio coding algorithms make use of asymmetrical window switching. This has the advantage of having simplified requirements of the transform windows for perfect reconstruction. Symmetrical window switching could also be used. One particular advantage of symmetrical window switching lies in the symmetry of the transform windows, which can be utilized by fast algorithms developed for performing the MLT.

8. CONCLUSION

The modulated lapped transform and its generalization the biorthogonal modulated lapped transform have been presented. An analysis of the conditions for perfectly reconstructing the BMLT has been performed and it has been found that contrary to previous findings time-domain aliasing cancellation of the transform window functions is not required for perfect reconstruction of neither the BMLT nor the MLT. Window switching in both asymmetrical and symmetrical fashion has been analyzed and the conditions of perfect reconstruction in each case have been given.

There are still some aspects worth further studies. For example the analysis of window switching could be performed in greater detail together with examples. It would in particular be interesting to study the behaviour of the symmetrical switching alternative more closely. One challenging area of investigation would be to consider window switching between transforms of different lengths. These remain as topics of future research.

9. REFERENCES

[1] H. Malvar, *Signal Processing with Lapped Transforms*. 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] ISO/IEC JTC1/SC29/WG11, “Information Technology - Very Low Bitrate Audio-Visual Coding, International Standard ISO/IEC 14496-3, Part 3: Audio,” 1998.

[6] S. Cheung and J. S. Lim, “Incorporation of biorthogonality into lapped transforms for audio compression,” in *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. V, (Detroit, MI), pp. 3079–3082, May 1995.

[7] J. Tuisku, “On the energy compaction performance of the biorthogonal modulated lapped transform,” in *Proceedings of the First IEEE Balkan Conference on Signal Processing, Communications, Circuits, and Systems*, (Maslak, Istanbul, Turkey), June 2000.

[8] J. Tuisku, “Transient performance of lapped transforms in audio coding,” in *Proceedings of the International Conference on Communications, Control, and Signal Processing*, (Bangalore, India), July 2000.

[9] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli, “Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms,” *IEEE Trans. on Signal Processing*, vol. 41, pp. 3341–3359, December 1993.

[10] S. Shlien, “The modulated lapped transform, its time-varying forms, and its applications to audio coding standards,” *IEEE Trans. on Speech and Audio Processing*, vol. 5, pp. 359–366, July 1997.

[11] M. Vetterli and D. Le Gall, “Perfect Reconstruction FIR Filter Banks: Some Properties and Factorizations,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1057–1071, July 1989.