

A Nonuniform Modulated Complex Lapped Transform

Zixiang Xiong and Henrique Malvar

July 2000

Technical Report
MSR-TR-2000-108

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
<http://www.research.microsoft.com>

A Nonuniform Modulated Complex Lapped Transform

Zixiang Xiong and Henrique Malvar

Abstract

A nonuniform modulated complex lapped transform (NMCLT) is introduced in this paper as a two-stage extension of the modulated complex lapped transform (MCLT). The NMCLT is a new nonuniform oversampled filter bank with a better combination of time- and frequency-domain localization than previous designs. Adaptive nonuniform subband decompositions can be easily generated by varying the number of coefficients brought to the second stage on a frame-by-frame basis. The NMCLT is ideally suited for processing wideband signals such as audio, with reduced time-spreading artifacts.

Index terms—Filter banks, modulated filter banks, lapped transforms, audio processing.

1 Introduction

The modulated lapped transform (MLT) [1] is as a kind of cosine-modulated filter bank [2] with the perfect reconstruction property. For applications such as audio coding, the MLT has two main advantages over block transforms such as the discrete cosine transform (DCT): no blocking artifacts and better rate-distortion performance [1]. That's why the MLT has been used in many audio coding systems, such as MPEG Layer III (MP3) and Dolby AC-3 [3].

A disadvantage of the MLT is that it introduces ringing artifacts (or pre-echo [3] in audio coding) at low bit rates because of the poor time resolution of the basis functions. One way to alleviate this problem is to switch to a shorter block length during high-frequency transient sounds, as it is done in MP3. Another way is to use hierarchical lapped transforms [1] or tree-structured wavelet packet [4] decompositions, both of which produce a nonuniform subband structure. This generates highpass basis functions with good time resolution but very poor frequency resolution due to aliasing. An alternative was introduced in [5, 6] to increase the time resolution of MLT basis functions by merging subbands such that basis functions with the same frequency resolution can have different time localizations. However, the number of subbands merged was very small (two or four) and no systematic way of merging subbands was offered in [5, 6]. Although a method similar to subband merging was used in [7, 8] to construct frequency-varying MLTs by cascading MLTs with smaller inverse MLTs, the resulting basis functions do not have as much time-domain separation as those in [5, 6].

In a related work, the modulated complex lapped transform (MCLT) was introduced in [9] as a simple extension to the MLT. A key observation made in [9] is that, with a $2\times$ oversampling ratio in the MCLT, time-domain aliasing terms in the real and imaginary reconstruction parts have opposite signs, i.e., they cancel each other.

In this paper, we introduce a nonuniform modulated complex lapped transform (NMCLT) by cascading a MCLT with shorter size MCLTs. Each shorter size MCLT plays the role of merging subbands. Time-aliasing terms in the NMCLT basis functions can be made to cancel each other in the inverse NMCLT. This allows us to construct basis functions (after aliasing cancellation) with desirable (e.g., wavelet packet-like) frequency resolutions and almost ideal time localizations. Fast algorithms for computing the modulated lapped sine and cosine transforms can be used to implement the NMCLT. Potential applications of the NMCLT include audio coding, denoising, and watermarking.

2 The Modulated Complex Lapped Transform

The basis functions of the MCLT are defined by cosine and sine modulation of the analysis window $h_a(n)$ and synthesis window $h_s(n)$ in the following form [9]:

$$p_a(n, k) = h_a(n)[\Phi^c(n, k) + j\Phi^s(n, k)]$$

$$p_s(n, k) = \frac{h_s(n)}{2}[\Phi^c(n, k) - j\Phi^s(n, k)]$$

with $j = \sqrt{-1}$, $0 \leq n \leq 2M - 1$, $0 \leq k \leq M - 1$, where M is the block size, and

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

$$\Phi^s(n, k) = \sqrt{\frac{2}{M}} \sin \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right].$$

For a real $2M$ -sample signal block \mathbf{x} , its corresponding vector \mathbf{X} of M MCLT coefficients $X(k)$ is computed as $\mathbf{X} = \mathbf{P}_a^T \mathbf{x}$, that is

$$X(k) = \sum_{n=0}^{2M-1} x(n)p_a(n, k), \quad 0 \leq k \leq M - 1. \quad (1)$$

The transform matrix $\mathbf{P}_a = \mathbf{C}_a + j\mathbf{S}_a$ is formed with $p_a(n, k)$ on its n -th row and k -th column. Likewise, the inverse transform matrix $\mathbf{P}_s = \mathbf{C}_s - j\mathbf{S}_s$ is formed with $p_s(n, k)$ as its entries. For a vector $\hat{\mathbf{X}}$ of processed MCLT coefficients, the reconstructed $2M$ -sample vector $\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = \text{Re}\{\mathbf{P}_s \hat{\mathbf{X}}\}$. Neighboring $\hat{\mathbf{x}}$ vectors are superimposed with M -sample overlap to form the reconstructed signal $\hat{x}(n)$ [9].

The analysis and synthesis windows are usually set to the MLT sine window [1]

$$h_a(n) = h_s(n) = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2M} \right] \quad (2)$$

for $0 \leq n \leq 2M - 1$, or they are related by (assuming the symmetry $h(n) = h(2M - 1 - n)$ for both windows)

$$h_a(n) = \frac{h_s(n)}{h_s^2(n) + h_s^2(M - 1 - n)}, \quad 0 \leq n \leq M - 1 \quad (3)$$

for a biorthogonal construction [5].

Clearly, the MLT of a signal is given by the real part of its MCLT. In addition, when there is no quantization or processing in the MCLT domain, i.e. $\hat{\mathbf{X}} = \mathbf{X}$, we have¹

$$\begin{aligned} \hat{\mathbf{x}} &= \text{Re}\{\mathbf{P}_s \hat{\mathbf{X}}\} = \text{Re}\{\mathbf{P}_s \mathbf{P}_a^T \mathbf{x}\} = \text{Re}\{\mathbf{P}_s \mathbf{P}_a^T\} \mathbf{x} \\ &= (\mathbf{C}_s \mathbf{C}_a^T + \mathbf{S}_s \mathbf{S}_a^T) \mathbf{x} = \text{diag}\{h_a(n)h_s(n)\} \mathbf{x}. \end{aligned} \quad (4)$$

The fact that $\text{Re}\{\mathbf{P}_s \mathbf{P}_a^T\}$ is diagonal is due to the $2 \times$ oversampling in the MCLT, which allows the time-domain aliasing terms in $\mathbf{C}_s \mathbf{C}_a^T$ and $\mathbf{S}_s \mathbf{S}_a^T$ to cancel each other *before* overlapping blocks of two neighboring $\hat{\mathbf{x}}$ vectors are superimposed. This fact suggests that it might be possible to cascade MCLTs of different sizes to construct basis functions with good time localizations (or no time-domain aliasing). This is indeed the main idea behind the NMCLT.

Finally, we note that

$$\text{Im}\{\mathbf{P}_s \mathbf{P}_a^T\} = \mathbf{S}_s \mathbf{C}_a^T - \mathbf{C}_s \mathbf{S}_a^T \neq \mathbf{0}. \quad (5)$$

This limits the input signal to the MCLT to be real. In addition, the MCLT reconstruction formula is not unique – one can also reconstruct the original signal from the only the real part or only the imaginary part of its MCLT [9].

3 The Nonuniform MCLT

Let the block size of the NMCLT be M . We construct the NMCLT from the MCLT, whose basis functions have uniform frequency resolution. Suppose we want to reduce the frequency resolution of the last rK MCLT basis functions by a factor of K , i.e., to improve the time localization of each of these basis functions by a factor of K . If we let $N = M - rK$, then what we want is to generate a $(N + r)$ -band nonuniform decomposition using the NMCLT, such that each of the first N bands has bandwidth π/M and each of the last r bands has bandwidth $K\pi/M$.

¹Throughout the paper, the superscript T when applied to a complex matrix denotes transpose (not conjugate transpose).

To generate such a nonuniform subband decomposition, we start with a signal vector \mathbf{x} of length $2M$, compute its corresponding MCLT vector \mathbf{X} of length M , and decompose it in two parts $\mathbf{X}^T = [\mathbf{X}_1^T \ \mathbf{X}_2^T]$ with $\mathbf{X}_1 = [\mathbf{X}(0), \dots, \mathbf{X}(N-1)]^T$ and $\mathbf{X}_2 = [\mathbf{X}(N), \dots, \mathbf{X}(M-1)]^T$. Then, we apply a second stage of MCLTs of length K to \mathbf{X}_2 . Note that the second-stage MCLTs are applied separately on the real and imaginary part of \mathbf{X}_2 because the input to an MCLT has to be real.

From this point on, we assume $h_a(n)$ and $h_s(n)$ are identical sine windows, as in (2). This allows us to drop the subscript in \mathbf{C}_a and \mathbf{S}_a and thus simplify the notation in the following description, which still holds when $h_a(n)$ and $h_s(n)$ are related according to (3).

Recall that $\mathbf{X} = \mathbf{P}_a^T \mathbf{x} = (\mathbf{C}^T + j\mathbf{S}^T)\mathbf{x}$. Also, let's partition \mathbf{C} and \mathbf{S} as $\mathbf{C} = [\mathbf{C}_1 \mid \mathbf{C}_2]$ and $\mathbf{S} = [\mathbf{S}_1 \mid \mathbf{S}_2]$, where the submatrices \mathbf{C}_1 and \mathbf{S}_1 are of size $2M \times N$, and \mathbf{C}_2 and \mathbf{S}_2 are of size $2M \times rK$. Then we can write $\mathbf{X}^T = \mathbf{x}^T [\mathbf{C}_1 + j\mathbf{S}_1 \mid \mathbf{C}_2 + j\mathbf{S}_2]$.

The second stage of MCLTs applied to $\text{Re}\{\mathbf{X}_2\} = \mathbf{C}_2^T \mathbf{x}$ and $\text{Im}\{\mathbf{X}_2\} = \mathbf{S}_2^T \mathbf{x}$ can be characterized in one single $rK \times rK$ transform matrix

$$\mathbf{Q} = \mathbf{C}_Q + j\mathbf{S}_Q = \begin{bmatrix} \mathbf{P}_1 & & & \mathbf{0} \\ & \mathbf{P}_0 & & \\ & & & \mathbf{P}_0 \\ \mathbf{0} & & & \mathbf{P}_2 \end{bmatrix}$$

where \mathbf{P}_0 is the $2K \times K$ transform matrix corresponding to a length- K MCLT. \mathbf{P}_1 and \mathbf{P}_2 are introduced because the first and last blocks of $\text{Re}\{\mathbf{X}_2\}$ and $\text{Im}\{\mathbf{X}_2\}$ only have one neighboring block [1].

Let the $M \times M$ matrix $\mathbf{U} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$ be an augmented version of \mathbf{Q} , where \mathbf{I}_N is the $N \times N$ identity matrix. Then, the two M -sample vectors of NMCLT coefficients corresponding to the $2M$ -sample signal block \mathbf{x} are given by

$$\begin{aligned} \mathbf{W}_a &= \mathbf{U}^T \text{Re}\{\mathbf{X}\} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^T \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T \end{bmatrix} \mathbf{x} \\ &= \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{Q}^T \mathbf{C}_2^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_Q^T \mathbf{C}_2^T + j\mathbf{S}_Q^T \mathbf{C}_2^T \end{bmatrix} \mathbf{x} \end{aligned} \tag{6}$$

and

$$\begin{aligned} \mathbf{W}_b &= \mathbf{U}^T \text{Im}\{\mathbf{X}\} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^T \end{bmatrix} \begin{bmatrix} \mathbf{S}_1^T \\ \mathbf{S}_2^T \end{bmatrix} \mathbf{x} \\ &= \begin{bmatrix} \mathbf{S}_1^T \\ \mathbf{Q}^T \mathbf{S}_2^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{S}_1^T \\ \mathbf{C}_Q^T \mathbf{S}_2^T - j\mathbf{S}_Q^T \mathbf{S}_2^T \end{bmatrix} \mathbf{x}. \end{aligned} \tag{7}$$

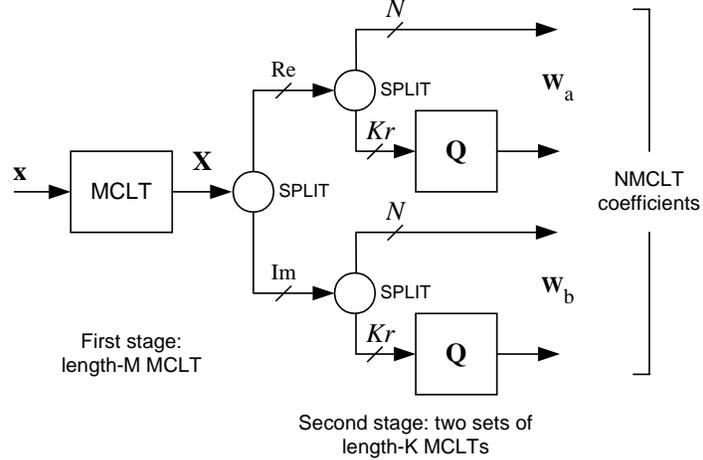


Figure 1: Simplified block diagram for the forward NMCLT. After a first MCLT stage producing complex coefficients in \mathbf{X} , two sets of length- K MCLTs are applied separately to the last Kr samples of the real and imaginary parts of \mathbf{X} .

Define $\mathbf{T}_a = [\mathbf{C}_1 \mid \mathbf{C}_2\mathbf{C}_Q + j\mathbf{C}_2\mathbf{S}_Q]$ and $\mathbf{T}_b = [\mathbf{S}_1 \mid \mathbf{S}_2\mathbf{C}_Q - j\mathbf{S}_2\mathbf{S}_Q]$ as the two forward NMCLT transform matrices. Then, the forward NMCLT transform is computed by

$$\mathbf{W}_a = \mathbf{T}_a^T \mathbf{x}; \quad \mathbf{W}_b = \mathbf{T}_b^T \mathbf{x}. \quad (8)$$

The steps above are depicted in Fig. 1.

An example of NMCLT basis functions $T_a(n, k)$ and $T_b(n, k)$ is shown in Fig. 2. As expected, the responses for \mathbf{T}_a and \mathbf{T}_b are quite similar. Each of them is a linear combination of basis functions in \mathbf{C}_2 or \mathbf{S}_2 , with weights coming from \mathbf{C}_Q and \mathbf{S}_Q . In other words, the NMCLT basis functions are generated by combining MCLT subbands via shorter size MCLTs, as shown in Fig. 1.

We see from (6)–(8) that the amount of oversampling in the NMCLT is larger than $2\times$ (as in the MCLT). This is because the first N elements of \mathbf{W}_a and \mathbf{W}_b are real, whereas their last Kr elements are complex. Thus, there are more options for NMCLT reconstruction than for MCLT reconstruction. We also see from Fig. 2 that there are still time-domain aliasing terms in the NMCLT basis functions (in fact, a bit more aliasing than in the NMLBT construction in [5]). However, by combining the basis functions of \mathbf{T}_a and \mathbf{T}_b as discussed below, the aliasing terms in the NMCLT basis functions can be effectively eliminated.

Let $\mathbf{W} = \mathbf{W}_a + j\mathbf{W}_b$ and $\mathbf{T} = \mathbf{T}_a + j\mathbf{T}_b$. Then, we see from (6) and (7) that $\mathbf{W} = \mathbf{U}^T \mathbf{X}$ and from (8) that $\mathbf{W} = \mathbf{T}^T \mathbf{x}$, with $\mathbf{T} = [\mathbf{C}_1 + j\mathbf{S}_1 \mid (\mathbf{C}_2\mathbf{C}_Q + \mathbf{S}_2\mathbf{S}_Q) + j(\mathbf{C}_2\mathbf{S}_Q - \mathbf{S}_2\mathbf{C}_Q)]$. The aliasing terms in $\mathbf{C}_2\mathbf{C}_Q$ and $\mathbf{S}_2\mathbf{S}_Q$ have opposite signs, while those in $\mathbf{C}_2\mathbf{S}_Q$ and $\mathbf{S}_2\mathbf{C}_Q$ have the same sign. In other words, the aliasing terms in $\text{Re}\{\mathbf{T}_a\}$ and $\text{Im}\{\mathbf{T}_b\}$ have opposite signs,

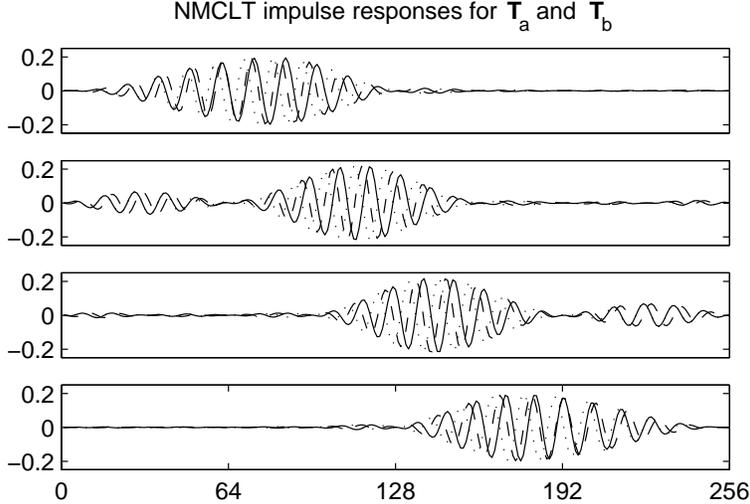


Figure 2: The NMCLT basis functions $T_a(n, k)$ and $T_b(n, k)$ for $21 \leq k \leq 24$, with $M = 128$, $N = 16$, and $K = 4$. Their real parts are plotted in solid lines, and the imaginary parts of T_a and T_b are plotted in dashed and dotted lines, respectively.

while those in $\text{Im}\{\mathbf{T}_a\}$ and $\text{Re}\{\mathbf{T}_b\}$ have the same sign. Therefore, all time-domain aliasing terms are cancelled. An example showing the better time-domain concentration of the basis functions in the combined transform matrix \mathbf{T} is shown in Figs. 3 and 4. We see from Fig. 4 that these new basis functions generated by the second stage have very good frequency domain characteristics; for each new set of K subbands, their passbands correspond to the sum of the K original MCLT subbands, without sacrificing stopband performance.

An inverse NMCLT transform can be computed by essentially reversing the steps in Fig. 1. If we call $\hat{\mathbf{W}}_a$ and $\hat{\mathbf{W}}_b$ the processed versions of the NMCLT transform coefficients, we can apply the inverse steps corresponding to second stage in Fig. 1 to obtain $\hat{\mathbf{X}}$, the processed version of $\hat{\mathbf{X}}$. Note that $\hat{\mathbf{X}}_1 = [\hat{\mathbf{W}}_a(0), \dots, \hat{\mathbf{W}}_a(N-1)]^T + j[\hat{\mathbf{W}}_b(0), \dots, \hat{\mathbf{W}}_b(N-1)]^T$. The reconstruction formula for $\hat{\mathbf{X}}_2$ is not unique. Based on the fact that the aliasing terms in $\text{Re}\{\mathbf{T}_a\}$ and $\text{Im}\{\mathbf{T}_b\}$ have opposite signs, we choose to compute $\text{Re}\{\hat{\mathbf{X}}_2\}$ from $\text{Im}\{[\hat{\mathbf{W}}_a(N), \dots, \hat{\mathbf{W}}_a(M-1)]^T\}$ via the inverse modulated lapped sine transform and $\text{Im}\{\hat{\mathbf{X}}_2\}$ from $\text{Re}\{[\hat{\mathbf{W}}_b(N), \dots, \hat{\mathbf{W}}_b(M-1)]^T\}$ via the inverse MLT. Finally, we generate the time-domain reconstructed signal $\hat{x}(n)$ by the MCLT inversion formula $\hat{\mathbf{x}} = \text{Re}\{\mathbf{P}_s \hat{\mathbf{X}}\}$.

Assuming equal quantization error levels in $\text{Re}\{\hat{\mathbf{W}}_b(k)\}$ and $\text{Im}\{\hat{\mathbf{W}}_b(k)\}$ for $N \leq k \leq M-1$, this choice of the NMCLT reconstruction formula guarantees that, after the cancellation of aliasing terms, the time-domain quantization error in $\hat{x}(n)$ will be a linear combination of basis functions in \mathbf{T} with almost ideal time localizations. This feature is very attractive in applications like compression and denoising, since the modification of high-frequency sub-

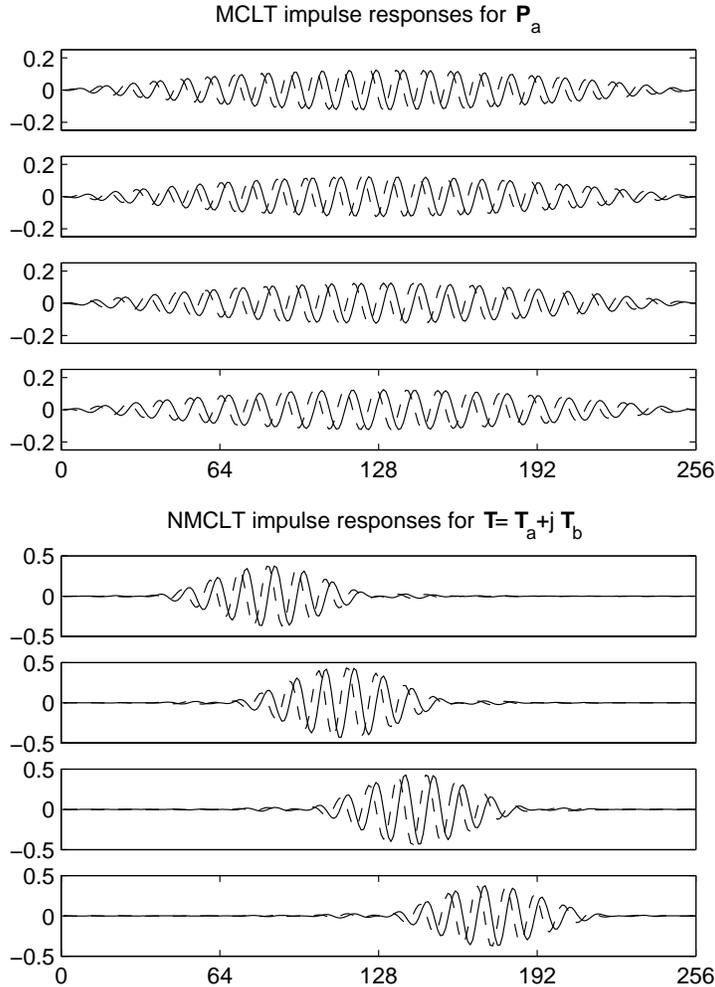


Figure 3: Time-domain responses of the original MCLT subbands $P_a(n, k)$ (top) and the corresponding new NMCLT subbands $T(n, k)$ (bottom) for $21 \leq k \leq 24$, with $M = 128$, $N = 16$, and $K = 4$.

bands will generate much shorter ringing, and therefore fewer reconstruction artifacts. For compression applications, the oversampling of the NMCLT is an issue; however, by appropriately taking into account the correlations among the NMCLT basis functions, we believe that rate-distortion performance of the NMCLT can be made very close to that of the MLT. This is currently a topic under investigation.

4 Conclusion

We have introduced the NMCLT as a new nonuniform oversampled filter bank design, by extending the MCLT with a second stage of shorter MCLTs applied to the high-frequency

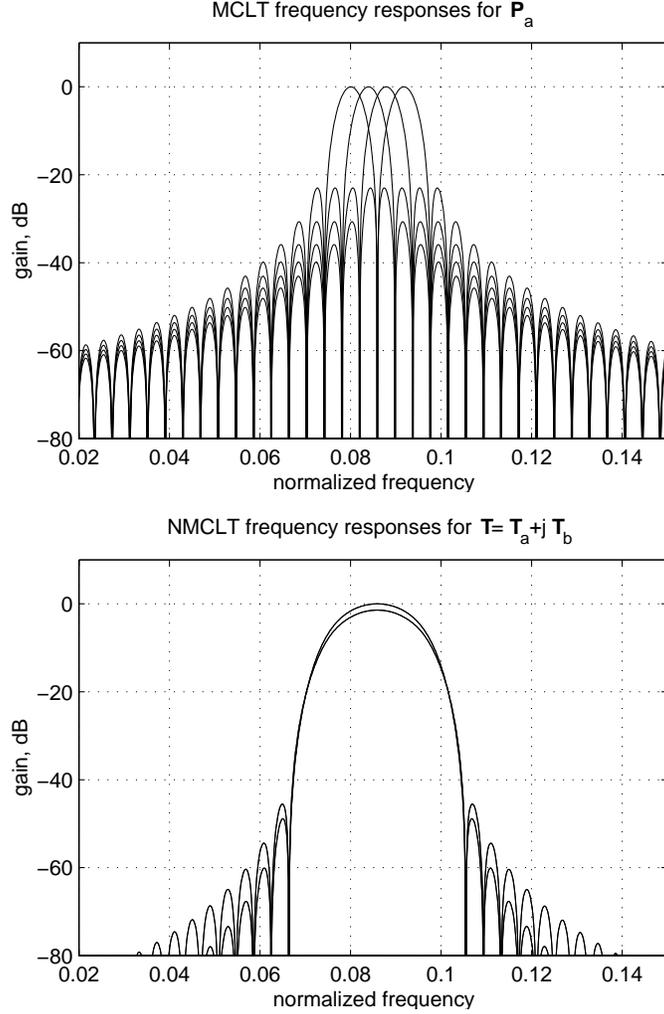


Figure 4: Frequency-domain responses of the original MCLT subbands $P_a(n, k)$ (top) and the corresponding new NMCLT subbands $T(n, k)$ (bottom) for $21 \leq k \leq 24$, with $M = 128$, $N = 16$, and $K = 4$.

subbands in a particular way. The resulting NMCLT has excellent time- and frequency-domain localization. Adaptive nonuniform subband decompositions can be easily generated by varying N on a frame-by-frame basis, which in many applications may be a better alternative than window-length switching. The NMCLT is ideally suited for processing wideband signals such as audio, in applications such as noise reduction and watermarking, with reduced time-spreading artifacts.

References

- [1] H. S. Malvar, *Signal Processing with Lapped Transforms*. Boston, MA: Artech House, 1992.
- [2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [3] S. Shlien, “The modulated lapped transform, its time-varying forms, and applications to audio coding,” *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 359–366, July 1997.
- [4] D. Sinha and A. Tewfik, “Low bit rate transparent audio compression using adapted wavelets,” *IEEE Trans. Signal Processing*, vol. 41, pp. 3463–3479, December 1993.
- [5] H. S. Malvar, “Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts,” *IEEE Trans. Signal Processing*, vol. 46, pp. 1043–1053, April 1998.
- [6] H. S. Malvar, “Enhancing the performance of subband audio coders for speech signals,” *Proc. ISCAS’98*, pp. 98–101, Monterey, CA, June 1998.
- [7] M. Purat and P. Noll, “Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transforms,” *Proc. ICASSP’96*, pp. 1021–1024, Atlanta, GA, May 1996.
- [8] R. L. de Queiroz, “Uniform filter banks with nonuniform bands: post-processing design,” *Proc. ICASSP’98*, pp. 1341–1344, Seattle, WA, May 1998.
- [9] H. S. Malvar, “A modulated complex lapped transform and its applications to audio processing,” *Proc. ICASSP’99*, pp. 1421–1424, Phoenix, AZ, March 1999 (a new version with typos corrected is available at <http://research.microsoft.com/~malvar/papers/icassp99.pdf>).