DISCRETE WAVELET TRANSFORM BASED SHIFT-INVARIANT ANALYSIS SCHEME FOR TRANSIENT SOUND SIGNALS

Wasim Ahmad, Hüseyin Hacıhabiboğlu†, and Ahmet M. Kondoz**

*I-Lab/CVSSP, University of Surrey, Guildford, GU2 7XH, United Kingdom
{w.ahmad, a.kondoz}@surrey.ac.uk
†CDSPR, King's College London, Strand, London, WC2R 2LS, United Kingdom
huseyin.hacihabiboglu@kcl.ac.uk

ABSTRACT

Discrete wavelet transform (DWT) has gained widespread recognition and popularity in signal processing due to its ability to underline and represent time-varying spectral properties of many transient and other nonstationary signals. However, DWT is a shift-variant transform. This shift-variance is a major problem with the use of DWT for transient signal analysis and pattern recognition applications. A number of modified forms of DWT have been investigated in recent years that provide approximate shift-invariant transform but at the cost of increased redundancy and complexity. In this paper, a shift-invariant analysis scheme is proposed which is nonredundant. This scheme combines minimum-phase (MP) reconstruction with the DWT so that the resultant scheme provides a shift-invariant transform. The detailed properties of MP signal and different methods to reconstruct it are explained. The proposed scheme can be used for the analysis-synthesis, classification, and compression of transient sound signals.

1. INTRODUCTION

A number of digital signal analysis techniques have been developed and applied to represent the transient sound signals. Discrete wavelet transform is the most suitable of these techniques because of its localization in both time and frequency [1], [2], and [3]. DWT provides an excellent framework for the analysis of transient sound signals as the redundancy and correlation among the resultant wavelet coefficients (WC) are very small. However, it is well known that DWT is a shift-variant transform which means the WC of two similar transient sound signals are considerably different even if the two signals just differ by time shift. The lack of shift-invariance in DWT causes problems when wavelet multiresolution representation is used in transient sound analysis, classification, identification, and detection applications [4], [5], [6], and [7].

In practice, it is highly desirable that when DWT is applied to a time-shifted signal, the time shift present in the signal should only shift the numerical descriptors of the WC rather than change them, otherwise recognition of the similar features could be complicated. To solve the shift-variance problem of DWT, a number of schemes have been proposed in recent years [8], [9], [10], and [11]. In these schemes, the wavelet coefficients are calculated with the fast filter bank algorithm [12] but without dyadic decimation. These schemes enable the restoration of the shift-invariance property of the DWT by modifying the conventional DWT, but they also increase redundancy, computational cost, and complexity of the transform. Therefore, additional compression and feature selection methods need to be used with these transforms for signal representation and in pattern recognition applications.

This paper presents a shift-invariant analysis scheme for finite-length transient sound signals that is based on minimum-phase reconstruction and discrete wavelet transform. In the proposed scheme, the MP signal is realized by decomposing the input signal into MP and all-pass (AP) components. The MP signals of the input signal and its time-shifted version are identical for suitably band-limited signal and therefore the DWT is applied to the MP version of the input signal. This shift-invariant analysis scheme is non-redundant, i.e., it maintains the compact representation of the signal and has a low computational complexity. The time shift present in the signal is extracted as an AP signal having the same phase as the original. The output signal can be reconstructed by reconstituting the phase from the AP signal and the processed MP signal. The presented shift-invariant analysis scheme can be potentially used in transient sound signals analysis-synthesis, morphing, detection, identification, and classification applications.

2. ANALYSIS OF TRANSIENT SOUND SIGNALS

2.1. Conventional Discrete Wavelet Transform

Discrete wavelet transform is computed using the Mallat's pyramidal algorithm, which is found to yield a fast computation of wavelet transform [12]. This algorithm uses the filter bank analysis to decompose the input signal into approximation and detail coefficients. Let $x[n]$ and $\overline{x}[n] = x[n+k]$ be the two finite length discrete-time transient sound signals of size N. The transient signal $\overline{x}[n]$ is a time-shifted version of $x[n]$ where $k \in \Re$ is the shift factor. The energy of the finite length signals $x[n]$ and $\overline{x}[n]$ can be calculated using Parseval's theorem as,

$$E = \sum_n |x[n]|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2 d\omega \quad (1)$$

$$\overline{E} = \sum_n |\overline{x}[n]|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\overline{X}(\omega)|^2 d\omega. \quad (2)$$

The signal $\overline{x}[n]$ is simply a time-shifted version of $x[n]$, therefore the total energy contained in $\overline{x}[n]$ is equal to the total energy of the original signal, i.e., $\overline{E} = E$. The Mallat's pyramidal algorithm is applied to the signals $x[n]$ and $\overline{x}[n]$ to decompose both the signals into approximation coefficients a_j, $\overline{a}_j$, and detail coefficients d_j, $\overline{d}_j$ where the subscript represents the level of decomposition. It can be shown that the total energy of each input signal is divided between the WC, i.e.,

$$E = \sum_n |x[n]|^2 = \sum_m |a_J[m]|^2 + \sum_{j=1}^{J} \sum_m |d_j[m]|^2 \quad (3)$$

$$\overline{E} = \sum_n |\overline{x}[n]|^2 = \sum_m |\overline{a}_J[m]|^2 + \sum_{j=1}^{J} \sum_m |\overline{d}_j[m]|^2 \quad (4)$$

where J is the highest number of decomposition. It can be observed from Eqs. (3) and (4) that DWT conserves the total energy of the input signals in their wavelet coefficients but because of its shift-variant nature, the WC of the signals $x[n]$ and $\overline{x}[n]$ are different at all decomposition levels, i.e., $a_j \neq \overline{a}_j$, and $d_j \neq \overline{d}_j$ for all $j = 1, 2, \ldots, J$. This phenomenon causes major variations in the distribution of energy in corresponding subbands of signals $x[n]$ and $\overline{x}[n]$, i.e.,

$$\sum_m |a_j[m]|^2 \neq \sum_m |\overline{a}_j[m]|^2 \quad (5)$$

$$\sum_m |d_j[m]|^2 \neq \sum_m |\overline{d}_j[m]|^2 \quad (6)$$

for all $j = 1, 2, \ldots, J$. The discrepancy between the WC and corresponding subbands energy pose many challenges in pattern recognition applications where the WC based features are used to train a model for the input signals.

2.2. Alignment of Transient Signals

When two similar transient signals such as $x[n]$ and $\overline{x}[n]$ are decomposed using DWT, the differences between their WC and subbands energy mainly come from the fact that they differ by a time shift. One approach to overcome or minimize this disparity consists in aligning the time-shifted input signals with the original signal. Let us take the Fourier transforms of original $x[n]$ and its time-shifted version $\overline{x}[n]$ transient signals, and express the magnitude and phase in polar form as,

$$X(\omega) = |X(\omega)| \angle X(\omega) \quad (7)$$

$$\overline{X}(\omega) = |X(\omega)| \angle (X(\omega) - \omega k) = X(\omega)\, e^{-j\omega k} \quad (8)$$

where k is the shift factor. When the shift factor is simply an integer i.e. $k \in \mathbb{Z}$, the alignment of such signals is a straightforward task. But in real time applications, any phase modification which involves a linear phase being added to signal causes a continuous shift in the signal, i.e. the shift is most likely to be a fractional number i.e. $k \in \mathbb{R}$. Secondly, there are situations where the phase shift is a nonlinear function of angular frequency ω, and consequently the shifted signal may look considerably different from the original. Furthermore, the two signals received from the same source at different time interval may look different because of variation in the added noise and/or external interference. Therefore, the alignment of such input signals is not a trivial task and can not be achieved easily.

3. PROPOSED SHIFT-INVARIANT ANALYSIS SCHEME

One of the key advantages of using the conventional DWT is that it provides a compact representation of a transient signal in time and frequency which can be computed efficiently. These properties of conventional DWT is compromised in the modified forms of DWT to achieve the shift-invariant transform [8], [9], [10], and [11]. We propose a scheme which does not compromise on these key advantages of the conventional DWT and serves as a shift-invariant analysis scheme for any finite-length transient signal.

The proposed shift-invariant analysis scheme consists of two stages. In the first stage, the MP signals are reconstructed from the input transient signals and in the second stage, DWT is applied to the MP version of the input signals. The purpose of MP reconstruction of the input transient signals is to remove the phase and any time shift present in them so that they become aligned. Therefore, the DWT is applied to the MP signals which are aligned and consequently produce the same set of WC and subbands energy at all decomposition levels. The block diagram of the proposed shift-invariant analysis scheme is depicted in Fig. 1.

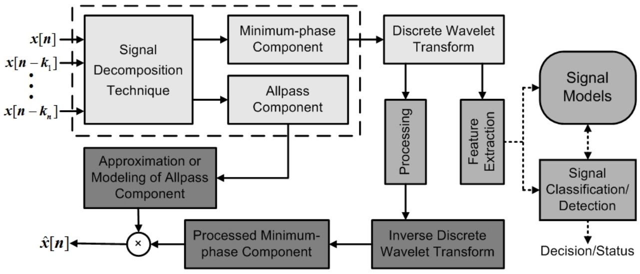


Figure 1: *DWT based shift-invariant analysis scheme.*

3.1. Minimum-Phase Signal

A signal is called minimum-phase if all zeros of its z-transform lie within the unit circle. When a signal is converted to its minimum-phase version, all the energy of the signal is transferred to the MP signal [13]. In other words, the magnitude spectrum of the input signal and its MP version are exactly the same [14]. An MP signal, also known as minimum-delay signal, starts at time $n = 0$ with maximum energy value and decays sharply with time. The magnitude and phase spectra of the MP signals, reconstructed from a finite length transient signal and its time-shifted versions, are exactly the same. Therefore, any finite length transient signal and its time-shifted version can be aligned by constructing their corresponding MP signals.

3.2. Construction of MP Signal

Any finite duration mixed-phase causal signal can be decomposed into minimum-phase and all-pass components where all the energy of input signal is extracted in the MP signal, and the phase and time shift present in the signal are presented as an AP signal. The decomposition can be obtained by either using a parametric method such as poles-zeros factorization or a nonparametric method such as cepstrum analysis [13], [14]. Let us take the transient signal $x[n]$ and its time-shifted version $\overline{x}[n]$ and decompose them into MP and AP signals using one of the method mentioned above. The signals can be represented in terms of the resultant components as,

$$x[n] = x_{mp}[n] * y[n] \quad (9)$$

$$\overline{x}[n] = \overline{x}_{mp}[n] * \overline{y}[n] \quad (10)$$

where $x_{mp}[n]$, $\overline{x}_{mp}[n]$ are MP signals and $y[n]$, $\overline{y}[n]$ are AP signals. The magnitude and phase spectra of Eqs. (9) and (10) can be written as;

$$X(\omega) = X_{mp}(\omega)\, Y(\omega) = (|X_{mp}(\omega)|\ \angle 0)\ (1\ \angle Y(\omega)) \quad (11)$$
$$\overline{X}(\omega) = \overline{X}_{mp}(\omega)\, \overline{Y}(\omega) = (|\overline{X}_{mp}(\omega)|\ \angle 0)\ (1\ \angle \overline{Y}(\omega)). \quad (12)$$

From Eqs. (7), (8), (11) and (12) that the $x_{mp}[n]$ and $\overline{x}_{mp}[n]$ have zero phase and same magnitude response, which implies that both MP signals are aligned and equal i.e.

$$X_{mp}(\omega) = \overline{X}_{mp}(\omega) \quad \Leftrightarrow \quad x_{mp}[n] = \overline{x}_{mp}[n]. \quad (13)$$

This is the case when $\overline{x}[n]$ is just a time-shifted version of $x[n]$. However, if the signals $x[n]$ and $\overline{x}[n]$ are two different samples from the same source, then the equality in the Eq. (13) becomes an approximation.

3.3. All-Pass Signal

The AP signal can be discarded, with reference to Fig. 1, when the proposed scheme is applied to extract WC based features which are used to model the classification or detection system. However, the AP is retained in the case where the input signal is reconstructed at the end by adding the phase from the AP signal to the MP signal. The AP signal can be either stored as it is or approximated by a finite impulse response (FIR) filter.

3.4. Shift-Invariant Decomposition

In the proposed scheme, DWT is applied to the MP version of the input signal. Therefore, the Mallat's pyramidal algorithm can be applied to signals $x[n]_{mp}$ and $\overline{x}[n]_{mp}$ which decomposes them into approximation coefficients $\breve{a}_j$, $\breve{\overline{a}}_j$ and detail coefficients $\breve{d}_j$, $\breve{\overline{d}}_j$. Consequently, this decomposition scheme produces the same set of approximation and detail coefficients at all decomposition levels, i.e., $\breve{a}_j = \breve{\overline{a}}_j$, and $\breve{d}_j = \breve{\overline{d}}_j$ for all $j = 1, 2, \ldots, J$. The presented scheme also produces the same energy distribution across corresponding subbands at all decomposition levels, i.e.,

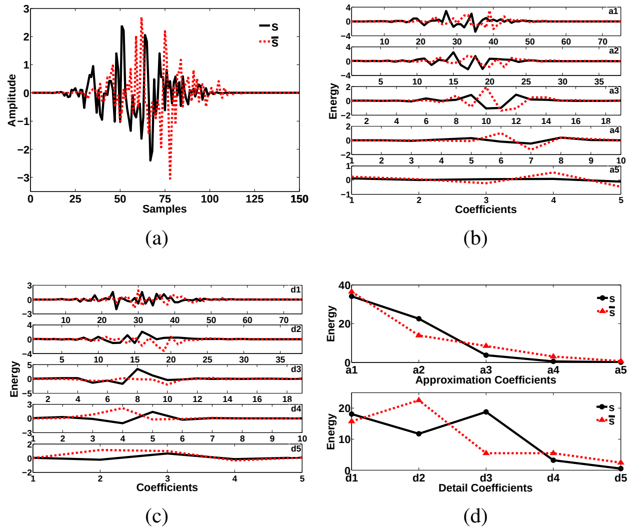
$$\sum_m |\breve{a}_j[m]|^2 = \sum_m |\breve{\overline{a}}_j[m]|^2 \quad (14)$$
$$\sum_m |\breve{d}_j[m]|^2 = \sum_m |\breve{\overline{d}}_j[m]|^2 \quad (15)$$

for all $j = 1, 2, \ldots, J$.

4. POTENTIAL APPLICATIONS

The proposed scheme is applicable to a number of applications, including analysis-synthesis, morphing, detection, identification, and classification of transient sound signals. In fact, the proposed shift-invariant scheme has already been employed by the author in the morphing of transient sounds where it was used to generate intermediate and other novel sounds [15]. It has potential usage in underwater sound detection, identification, and classification applications where sound features and their energy are sensitive to time shift.

To illustrate the effectiveness of the proposed shift-invariant analysis scheme, it is applied to a synthetic transient signal and its time-shifted version. A synthetic transient signal s was generated from normally distributed random numbers enveloped by a symmetrical Hann window. Another transient signal $\overline{s}$ was obtained by shifting the generated signal s by 10.7 samples. The rational delay in the signal s was realized using a Thiran fractional-delay filter [16]. The transient signals s and $\overline{s}$ were analyzed using DWT up to 5^{th} decomposition level. The input signals and their decomposition coefficients are plotted in Fig. 2a, 2b, and 2c respectively. It can be observed from Fig. 2b, and 2c that the WC of the signals s and $\overline{s}$ are not identical at each decomposition level. The total energy in each frequency band is also plotted in Fig. 2d. It can also be observed that the energy distribution in corresponding subbands is different for the original signal and its time-shifted version.



(a) (b) (c) (d)

Figure 2: *(a) The original signal s and its delayed version $\overline{s}$, (b) approximation coefficients a_j & $\overline{a}_j$, (c) detail coefficients d_j & $\overline{d}_j$, (d) energy distribution in subbands.*

The proposed scheme decomposes the signals s and $\overline{s}$ into MP components s_{mp}, $\overline{s}_{mp}$, and AP components y and $\overline{y}$. It can be observed from Fig. 3a that the MP signals s_{mp} and $\overline{s}_{mp}$ are similar and aligned. Now the application of DWT on extracted MP signals of s and $\overline{s}$ generates the same set of approximation and detail coefficients, as depicted in Figs. 3b, and 3c. The energy distribution in corresponding subbands is also identical for MP versions of s and $\overline{s}$ at all decomposition levels as shown in Fig. 3d. It can also be observed from the magnitude and phase responses of MP and AP components, Fig. 3e, that the phase of input signal and any time shift present are extracted in AP signals. The time shift present in $\overline{s}$ appeared as additional delay in the phase response $\overline{y}$ as compared to y.

5. CONCLUSIONS AND FURTHER RESEARCH

A DWT based shift-invariant transform was presented in this paper, which can be used to represent a signal in analysis-synthesis and compression applications, or to extract features in signal detection and classification applications. The proposed scheme is a shift invariant in a sense that it produces the same set of WC and subbands energy at all decomposition levels regardless of any time shift present in the input signals. This scheme maintains the properties of conventional DWT such as compact representation and computational efficiency.

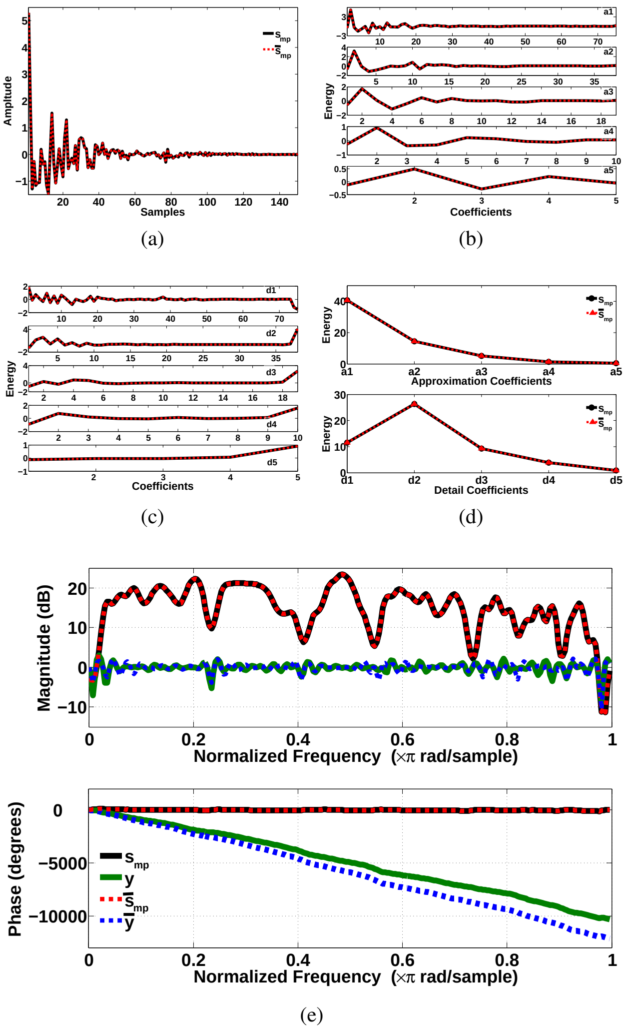


Figure 3: *(a) MP sequences of original signal s_{mp} and its delayed version $\overline{s}_{mp}$, (b) approximation coefficients $\breve{a}_j$ & $\breve{\overline{a}}_j$, (c) detail coefficients $\breve{d}_j$ & $\breve{\overline{d}}_j$, (d) energy distribution in subbands, (e) the magnitude and phase response of MP and AP components.*

The input transient signals were aligned by constructing their MP versions such that each input signal and its MP version contain the same energy. The DWT was applied to the MP version of input signal. The entire phase and any time shift present in each input signal were extracted as AP signal. The proposed scheme was applied on a synthetic signal which produced the same set of WC and subbands energy at all decomposition levels.

As part of future work, we will compare different methods to construct MP and AP components and their real time implementation as well as investigate different ways to approximate the AP signal using either FIR filters or any other meaningful compact representation. It would also be interesting to investigate how the construction of MP signal will affect the characteristics of the attack part of the input transient signal as it might play an important role in sound identification and classification.

6. REFERENCES

[1] R. Kronland-Martinet, "The wavelet transform for analysis, synthesis, and processing of speech and music sounds," *Comp. Music J.*, vol. 12, no. 4, pp. 11–20, 1988.

[2] G. Tzanetakis, G. Essl, and P. Cook, "Audio analysis using the discrete wavelet transform," in *Proc. of WSES Int. Conf. on Acoust. and Music: Theory and Appl.*, Greece, 2001.

[3] S. Dubnov, Z. Bar-Joseph, R. El-Yaniv, D. Lischinski, and M. Werman, "Synthesizing sound textures through wavelet tree learning," *IEEE Comput. Graph.*, vol. 22, no. 4, pp. 38 – 48, July 2002.

[4] J. Liang and T. W. Parks, "A translation-invariant wavelet representation algorithm with applications," *IEEE T. Signal. Proces.*, vol. 44, no. 2, pp. 225–232, Feb 1996.

[5] O. Duzenli and M. P. Fargues, "Wavelet-based feature extraction methods for classification applications," in *Proc. 9th IEEE SP Workshop on Stat. Signal and Array Proc.*, 1998, pp. 176 – 179.

[6] Huilin Xiong, Tianxu Zhang, and Y. S. Moon, "A translation- and scale-invariant adaptive wavelet transform," *IEEE T. Image. Process.*, vol. p, no. 12, pp. 2100–2108, Dec 2000.

[7] F. T. Wang, J. C. Y. Lee, S. H. Chang, C. P. Chou, H. H. Chang, and Y. H. Wang, "Signal detection in underwater sound by dual-tree discrete wavelet transform," in *Proc. OCEANS 2007 - Europe*, Jun 2007, pp. 1–4.

[8] S. G. Mallat, "Zero-crossing of a wavelet transform," *IEEE T. Inform. Theory.*, vol. 37, no. 4, pp. 1019–1033, July 1991.

[9] G. P. Nason and B.W. Silverman, "The stationary wavelet transform and some statistical applications," in *Lec. Notes in Stat., 103*, pp. 281–299. 1995.

[10] R. R. Coifman and D. L. Donoho, "Translation invariant denoising," in *Lec. Notes in Stat., 103*, pp. 125–150. 1995.

[11] J. C. Pesquet, H. Krim, and H. Carfatan, "Time-invariant orthonormal wavelet representations," *IEEE T. Signal. Proces.*, vol. 44, no. 8, pp. 1964–1970, Aug 1996.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[13] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall International Inc., int. edition, 1989.

[14] A. V. Oppenheim, G. E. Kopec, and J. M. Tribolet, "Signal analysis by homomorphic prediction," *IEEE T. Acoust. Speech.*, vol. 24, no. 4, pp. 327 – 332, Aug 1976.

[15] W. Ahmad, H. Hacıhabiboğlu, and A. M. Kondoz, "Morphing of transient sounds based on shift-invariant discrete wavelet transform and singular value decomposition," in *Proc. of IEEE Int. Conf. on Acoust., Speech, and Signal Proc.*, Taipei, Taiwan, Apr 2009, pp. 297–300.

[16] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay-tools for fractional delay filter design," *IEEE Signal. Proc. Mag.*, vol. 13, no. 1, pp. 30–60, 1996.