Application of The Wavelet Transform In The Processing of Musical Signals

Group Members:
Anshul Saxena  anshuls@ee.iitb.ac.in  01d07027
Sanjay Kumar  skumar@ee.iitb.ac.in  01d07041
Rajesh Meena  rajeshm@cse.iitb.ac.in  01d05016

Abstract

Wavelet has found a number of applications in the field of music. In this report some of these applications are explored. The use of wavelet in obtaining signal transformation, equalization, pitch shifting and pitch detection is considered. Finally an algorithm to denoise musical signal is obtained.

I. INTRODUCTION

The wavelet transform decomposes a signal into linear combination of basis functions which are all derived from one mother wavelet by means of dilation(scale) and translation(time). The transform on one particular scale is a bandpass filter since a wavelet is localized in frequency. On all scales these filters have the same relative bandwidth since all wavelets are derived by dilation from the mother wavelet. The time-frequency analysis thus studies low frequency with more frequency detail but less time resolution than high frequencies, which get a better time resolution. Music is also a typical time-frequency phenomenon. The notes contain frequency information(pitch) and time information (duration, starting time). The frequency information is logarithmically divided: raising one octave doubles the frequency. It is thus necessary to analyse musical signals with more frequency detail for the low frequencies and less frequency detail for the high frequencies. First section includes the background theory of wavelet transform, next section includes processing of these signals using many different wavelet.

II. BACKGROUND THEORY

A. Wavelet Transformation

In the continuous wavelet transform, we decompose a signal $x(t)$ into a linear combination of basis functions by the following formula which is known as the analysis formula.

$$W_{\psi}x(b,a) = \frac{1}{|a|} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt$$  \hspace{1cm} (1)

We can recover the signal $x(t)$ from the wavelet coefficients using the synthesis coefficient

$$x(t) = \frac{1}{c_{\psi}} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \frac{1}{|a|} W_{\psi}x(b,a)\psi\left(\frac{t-b}{a}\right)da \, db$$  \hspace{1cm} (2)
Now, the analysis wavelet can be changed in the reconstruction, allowing some aspects of the signal to be preserved while changing others. While simply changing the wavelet to be used in the inverse transform is a straightforward transformation, in practice the choice of the wavelet function to be used can be critical. For example, if one uses Mallat’s smooth wavelet for the forward transform and the Haar’s boxcar wavelet for the inverse, the result is that a great deal of noise seems to have been added to the signal, since the difference between these wavelets is considerable.

**B. Wavelet Equalisation**

A second possibility is to change the values of the wavelet domain coefficients; this acts much as an equalizer, changing the behaviour of the signal at a certain frequency level. Thus, for example, we can start with a signal which is very rich in spectral content and than remove certain frequency ranges to leave a different and more interesting sound. Since wavelet representation involve both translation and scale parameters, it becomes fairly simple to impose amplitude envelopes at different scales, so that the frequency content of the signal could be made to change with time.

**III. PROCESSING MUSICAL SIGNAL**

**A. Pitch Shifting/ Time Stretching**

One interesting possibility offered by wavelet transform is to change the pitch of the signal without changing the duration or conversely changing the duration of the signal without changing the pitch of the signal. Apart of musical usage of such techniques, there are useful application in a number of fields.

The method decomposes an audio signal in its CWT using the Morlet wavelet, changes the scale-axis, and retransforms the CWT to a signal. Changing the scale-axis is obviously the tricky part. If one wants to raise the pitch of an audio signal by a factor c, and therefore divide all the scales in the CWT of the signal by c, and retransforms the result, one does not get the expected result. The point is that one can not just change the coefficients of the CWT. Hence modification of these coefficients must be done with care. This is achieved using complex morlet wavelet. The phase of the coefficient is related to the frequency of the signal analyzed at the scale of the coefficients considered, so if we divide the scales by a factor c, and change the phases of the coefficients accordingly, we get the desired result. The procedure is shown in fig 2.

**B. Pitch Detection of Musical Signals**

Pitch period is a fundamental parameter in the analysis process of any physical model. A pitch detector is basically an algorithm that determines the fundamental pitch period of an input musical signal. Pitch detection algorithms can be divided into two groups: time-domain pitch detectors and frequency-domain pitch detectors. Pitch detection of musical signals is not a trivial task due to some difficulties such as the attack transients, low frequencies, and high frequencies.
1) Autocorrelation method: The autocorrelation function is a time-domain pitch detector. It is a measure of similarity between a signal and translated (shifted) version of itself. The basic idea of this function is that periodicity of the input signal implies periodicity of the autocorrelation function and vice versa.

For non-stationary signals, short-time autocorrelation function for signal $f(n)$ is defined below

$$ ph_l(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} [f(n+l)w(n+l)][f(n+m+l)w(n+m+l)] $$

(3)

$$ 0 < m < M_0 - 1 $$

where $w(n)$ is an appropriate window function, $N$ is the frame size, $l$ is the index of the starting frame, $m$ is the autocorrelation parameter or time lag and $M_0$ is the total number of points to be computed in the autocorrelation function. The autocorrelation function has its highest peak at $m=0$ which equals to the
average power of the input signal. For each l, one searches for the local maxima in a meaningful range of m. The distance between two consecutive maxima is the pitch period of the input signal \( f(n) \). Different window functions such as rectangular, Hanning, Hamming, and Blackman windows have been used in the analysis. The choice of an analysis window and the frame size are among the main disadvantages of the autocorrelation function.

2) Dyadic Wavelet Transform Method: Wavelet transform is based on the idea of filtering a signal \( f(t) \) with a dilated and translated versions of a prototype function \( \psi(t) \).

Dyadic Wavelet Transform (DWT), is the special case of CWT when the scale parameter is discretized along the dyadic grid \( 2^j \), \( j=1,2... \) and \( b \in \mathbb{Z} \).

\[
DWT(f, j) = W_j f = f(t) \ast \psi_{2^j}(t)
\]

where \( \ast \) denotes convolution and

\[
\psi_{2^j}(t) = \frac{1}{2^j} \psi\left(\frac{t}{2^j}\right)
\]

For an appropriately chosen wavelet, the wavelet transform modulus maxima denote the points of sharp variations of the signal. This property of DWT has been proven very useful for detecting pitch periods of speech signals[3]. An appropriately chosen wavelet is a wavelet that is the first derivative of a smooth function. Zero-crossings of musical signals can be considered as points of sharp variation of the signal and hence the dyadic wavelet transform exhibits local maxima at these points across several consecutive scales. The pitch period is evaluated by measuring the time distance between two such consecutive maxima. Spline wavelet is used for pitch detection as it is the first derivative of a smooth function.

C. Denoising Musical Signals Using Wavelet Bases

Audio denoising by investigating whether wavelet packet and local trigonometric packet bases can be used to successfully decompose the signal for processing. The method on which the algorithm is based around is to decompose a window of a signal using a wavelet packet or local trigonometric packet transform into a full binary tree of bases. Using an entropy measure, the tree is “pruned” to obtain a complete, non-redundant representation of the signal. Most audio signals are far too long to be processed in their entirety. Thus, it is necessary to divide the time-domain signal into windowed intervals and process each window individually.

- First, the window length must be chosen. Windows which are too short fail to pick up the important time structures of the audio signal. In addition to choosing the length of the window, an appropriate windowing function has to be determined. We use overlapping windows. That is, each window shares some samples in common with its neighbours.
- In step two of the algorithm, the windowed signal is decomposed in each basis of a collection of bases, called a basis “library”. Bases include wavelet packet bases constructed from different kinds of wavelets.
- For each basis tree in the library, the entropy is calculated for each node, and the tree is pruned to find the best basis within that tree. The entire tree is then given an entropy measure by adding together the entropies of the lowermost nodes remaining on the pruned tree. Selects the tree from the library with the lowest total entropy.
- Decompose the windowed signal in that wavelet basis.
- Determine which coefficients form a specific part of the underlying signal, and which coefficients can be discounted as noise. Coefficients corresponding to the signal are dubbed the coherent coefficients.
- Discard the coefficients which are not coherent.
- Transform the packet coefficients back to the time-domain, to form a denoised version of the original signal window.
- Merges the denoised window in with the rest of the reconstructed windows, performing the crossfading alluded to above if there is any overlap with adjacent windows.
Fig. 3. Best Basis Denoising Procedure

ACKNOWLEDGMENT

The authors would like to thank Prof. V.M.Gadre for his consistent support and guidance.

REFERENCES