A WAVELET-BASED PITCH DETECTOR FOR MUSICAL SIGNALS

John Fitch
Department of Mathematical Sciences,
University of Bath,
Bath BA2 7AY, UK
e-mail: jpff@maths.bath.ac.uk

Wafaa Shabana
Department of Mathematical Sciences,
University of Bath,
Bath BA2 7AY, UK
e-mail: wrs@maths.bath.ac.uk

ABSTRACT

Physical modelling of musical instruments is one possible approach to digital sound synthesis techniques. By the term physical modelling, we refer to the simulation of sound production mechanism of a musical instrument, which is modelled with reference to the physics using wave-guides. One of the fundamental parameters of such a physical model is the pitch, and so pitch period estimation is one of the first tasks of any analysis of such a model.

In this paper, an algorithm based on the Dyadic Wavelet Transform has been investigated for pitch detection of musical signals. The wavelet transform is simply the convolution of a signal f(t) with a dialated and translated version of a single function called the mother wavelet that has to satisfy certain requirements. There are a wide variety of possible wavelets, but not all are appropriate for pitch detection. The performance of both linear phase wavelets (Haar, Morlet, and the spline wavelet) and minimum phase wavelets (Daubechies' wavelets) have been investigated. The algorithm proposed here has proved to be simple, accurate, and robust to noise; it also has the potential of acceptable speed. A comparative study between this algorithm and the well-known autocorrelation function is also given. Finally, illustrative examples of different real guitar tones and other sound signals are given using the proposed algorithm.

KEYWORDS

Physical modeling – wavelet transform – pitch – autocorrelation function.

1. INTRODUCTION

During the last two decades, physical modelling of real musical instruments has gained popularity as a tool for sound synthesis and computer music. The term physical modelling refers to the simulation of sound production mechanism and the behaviour of a real musical instrument [1] [2] [3].

In physical modelling of a guitar (as a plucked string instrument), the ideal vibrating string is considered as the main source of vibration. It satisfies the one-dimensional wave equation, which can be modelled very accurately using digital wave-guide techniques [4]. Starting with a recorded real guitar tone, estimating the model parameters is one of the main tasks of the analysis process. Hence, pitch period estimation is essential for extracting the other parameters. Unlike speech signals, musical signals have a broader range of frequencies, so there are some difficulties in estimating their pitch period [5]. The autocorrelation function is one of the well-known time-domain pitch detectors.

Despite its simplicity, the autocorrelation function has some disadvantages.

An algorithm based on the dyadic wavelet transform has been investigated for pitch estimation of musical signals. The basic idea of this algorithm is that, for an appropriately chosen wavelet, the dyadic wavelet transform exhibits local maxima at the points of sharp variation of the signal [6].

In this paper, the application of the proposed algorithm to a wide range of stringed musical signals as well as some other musical signals has been investigated. Further, a comparative study between this algorithm and the autocorrelation function is presented. This paper is organized as follows: section 2 is devoted to the pitch detection problem and the autocorrelation algorithm. In section 3, principles of the dyadic wavelet transform and its properties is presented. In section 4, implementation of the proposed algorithm and the autocorrelation algorithm to a wide range of musical signals as well as singing voices is studied. Discussions and results are presented in section 5. Finally, section 6 is devoted to the conclusion.

2. 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.

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 as [7]:

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

$$0 \le m \le M_0 - 1,$$

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.

3. DYADIC WAVELET TRANSFORM

Wavelet transform is based on the idea of filtering a signal *f*(t) with a dialated and translated versions of a prototype function $\Psi(t)$. This function is called the mother wavelet and it has to satisfy certain requirements [8]. The *Continuous Wavelet Transform* (CWT) for *f*(t), is defined as [9];

$$CWT(f,a,b) = \int_{-\infty}^{\infty} f(t)\Psi_{a,b}(t)dt\,, \qquad (2)$$

where $\Psi_{a,b}(t) = \Psi(\frac{t-b}{a})$, a$\in R-\{0\}$ is the scale parameter and b$\in R$ is the translation parameter. In addition to its simple interpretation, the CWT satisfies some other useful properties such as linearity and conservation of energy [8] [9]. For practical implementations, CWT is computationally very complex.
*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$\in Z$ [10], i.e.,

$$\text{DWT(f,j)}= W_j f = f(t) * \Psi_{2^j}(t) \qquad (3)$$

where * denotes convolution and $\Psi_{2^j}(t) = \frac{1}{2^j}\Psi(\frac{t}{2^j})$. For an appropriately chosen wavelet, the wavelet transform modulus maxima denote the points of sharp variations of the signal [6] [10] [12]. This property of DWT has been proven very useful for detecting pitch periods of speech signals [11]. An appropriately chosen wavelet is a wavelet that is the first derivative of a smooth function [6]. 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.

4. IMPLEMENTATION

Theoretically, the *Dyadic wavelet transform* has to be evaluated for all scales (2^j), for j varying from $-\infty$ to $+\infty$. For practical implementation, one is limited to a finite larger scale and a nonzero finer scale, since the input signal is generally measured with a finite resolution. The finer scale is equal to 1 (for normalization purposes) and the larger scale is equal to 2^J. The wavelet used in this analysis is the quadratic spline wavelet, which is the first derivative of the cubic spline $\theta(t)$, i.e.,

$$\hat{\Psi}(\omega) = i\omega\hat{\theta}(\omega), \qquad (4)$$

with

$$\hat{\theta}(\omega) = \exp(-i\omega)\left(\frac{\sin(\omega/4)}{(\omega/4)}\right)^4 \qquad (5)$$

This wavelet is an anti-symmetric, regular and of compact support. The corresponding scale function $\Phi(t)$ is the quadratic spline with Fourier transform given by

$$\hat{\Phi}(\omega) = \exp(-i\frac{3\omega}{2})\left(\frac{\sin(\omega/2)}{(\omega/2)}\right)^3 \qquad (6)$$

Figure 1. show both $\Psi(t)$ and $\Phi(t)$ respectively.
Two FIR filters, namely, a low-pass filter $\{h(\text{n})\}$ and a high-pass filter $\{g(\text{n})\}$ characterize the discrete dyadic wavelet transform, and the number of levels J. Starting with $S_0 f = f$, $h_0 = h$ and $g_0 = g$, the recursive algorithm is defined as

$$\left.\begin{aligned} W_{j+1} f &= g_j * S_j f \\ S_{j+1} f &= h_j * S_j f \end{aligned}\right\}, \qquad (7)$$

j=0,1…J-1.
where h_j and g_j denotes the filters obtained from h and g by inserting $2^j - 1$ zeros between each two consecutive coefficients of the two filters respectively. Hence the DWT can be implemented as a FIR non-subsampled octave-band filter bank.

(a) quadratic spline wavelet (b) quadratic spline function

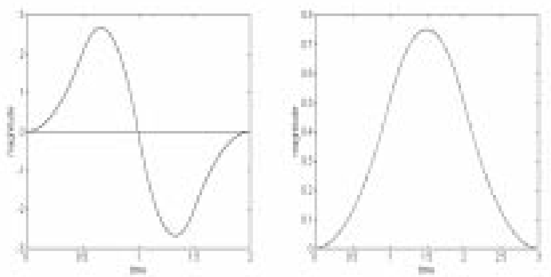

Figure (1)

5. RESULTS AND DISCUSSONS

The proposed algorithm has been implemented on a wide range of musical signals such as a saxophone signal (wind instrument), a tanpura signal (an Indian drone instrument), a singing voice signal, and a conga rim-hit signal (drum family) but with emphasis on plucked string signals (a classical guitar tone, bass, pizzacato cello ...etc).
The sampling rate for all test signals is 44.1 kHz and different window size of 22.7 and 34 ms has been used. Experiments have shown that it is adequate to evaluate the dyadic wavelet transform across three consecutive scales only 2^4, 2^5, and 2^6.
In the analysis of plucked string signals, the results show that it is sufficient to estimate the pitch period from the steadily decaying part of the signal several hundred milliseconds after the attack [13]. This is due to the fact that pitch period of plucked-string signal decreases as the signal attenuates. The test signal is a D-tone guitar and the estimated pitch period is 147 Hz. In this case, the relative error is (0.001). Results for guitar tone is shown in figure (2).

The estimated pitch of a D#-sax signal is 154.7368 Hz with relative error 0.005. Figure (3) shows results of the sax signal.

The tanpura signal is a very harmonically rich signal. Our results show that the proposed algorithm has the ability to detect not only the fundamental frequency but also the frequency with the most energy present in the signal. In the case of the autocorrelation function, the effect of the window function is to taper the function smoothly to 0. Hence, a longer frame size has to be used in order to detect the fundamental of the tanpura signal not the strongest harmonic. The estimated pitch period of this signal is 157.5 Hz. Results of this signal are shown in figure (4).

Figure (5) shows results of the analysis of a male singing voice. The estimated pitch is 110.8040 Hz.

Moreover, in the analysis of conga-rim signal, the algorithm classified this signal as unpitched one since it failed to find local maxima that satisfy the previous criteria.

For all test signals, the results can be further improved by using several methods of curve fitting for best estimate of local maxima.

The computational complexity of the proposed algorithm is O(NJ), for an input signal of length N evaluated across J scale. The constant depends on the number of the nonzero coefficients present in the filters *h* and *g*. The algorithm is faster than the autocorrelation function since the length of the analysis wavelet is less than M_0.

Different wavelets like Haar wavelet [14], a minimum-phase wavelet [14], and Morlet wavelet [15], have been used in the analysis to compare their performance. The spline wavelet has a superior performance. Results also show that the Haar wavelet has the potential of real-time implementation due to its simplicity and its accurate results.

Despite its simplicity, the autocorrelation function is computationally expensive when the appropriate frame size is used. Its main drawback is the choice of a window function and assuming stationarity of the signal within the frame, hence using a fixed frame size during the analysis process. More about the analysis process of all test signals is found in [16].

6. CONCLUSIONS

An algorithm based on the Dyadic Wavelet Transform is investigated for pitch detection of musical signals. The results show that the algorithm can be applied to a wide range of musical signals such as guitar, sax, cello, bass, tanpura as well as some singing voices. The algorithm is simple since only two FIR-filters are required for the analysis. It is accurate, efficient and robust to noise. The main advantage of the proposed algorithm is that it is fast compared to the autocorrelation function. Besides, the algorithm takes into account the non-stationarity of the input signal. Unlike the autocorrelation function, the frame size in not a crucial parameter since different frame sizes have been used successfully. On comparing the performance of different wavelets, the quadratic spline wavelet has a superior performance. Nevertheless, the algorithm has the potential of real-time implementation using the Haar wavelet due to its simplicity with minimal loss of accuracy. Finally, the algorithm can classify unpitched signals.

REFERENCES

[1] Valimaki, V. and Takala, T., "Virtual Musical Instruments-natural sounds using physical models", Organized Sound, 1(2): 75-86, (1996).

[2] Valimaki, V., Huopaniemi, J., Karjalainen, M. and Janosy, Z., "Physical modeling of plucked string instruments with application to real-time sound synthesis", Journal of the Audio Engineering Society, 44(5): 331-353, (1996).

[3] Karjalainen, M., Valimaki, V. and Tolonen, T., "Plucked string models-from Karplus-strong algorithm to digital waveguides and beyond", Computer Music Journal, 22(3): 17-32, (1998).

[4] Smith, J.O., " Physical modeling using digital waveguides", Computer Music Journal, 16(4):74 -91, (1992).

[5] Roads, C., " The computer music tutorial " , The MIT Press, Cambridge, Massachusetts, USA, 504-520, (1995).

[6] Mallat S. and Zhong S., " Characterization of signals from multiscale edges", IEEE Trans. on Pattern Analysis and Machine intelligence, 14(7): 710-732, (1992).

[7] Rabinar L R. , " On the use of autocorrelation analysis for pitch detection", IEEE Trans. on Acoustics, Speech and Signal Processing, 25(1):24- 33, (1977).

[8] Kronland Martinet, R. , Morlet, J. and Grossman, A., "Analysis of sound patterns through wavelet transforms", International Journal of Pattern Recognition and Artificial Intelligence, 1(2): 273-302, (1987).

[9] Sadowsky, J., "The continuous wavelet transform: a tool for signal investigation and understanding", Johns Hopkins APL Technical Digest, 15(4): 306-318, (1994).

[10] Mallat, S., " Zero- crossing of a wavelet transform " , IEEE Trans. on Information Theory, 37(4): 1019-1033, (1991).

[11] Kadambe, S., and Boudreaux-Bartels, G., "Application of the wavelet transform for pitch detection of speech signals", IEEE Trans. on Information Theory, 38(2): 917-924, (1992).

[12]Berman, Z. and Baras, J. S., " Properties of the multiscale maxima and zero-crossings representations", IEEE Trans.on Signal Processing, 41(12): 3216-3231, (1993).

[13]Tolonen, T., " Model-based analysis and resynthesis of Acoustic Guitar Tones", Helsinki University of Technology Department of Electerical and Communications Engineering, Technical report 46, (1998).

[14]Daubechies, I., "Orthonormal bases of compactly supported wavelets", Commun.Pure and Applied Math, 41: 909-996, (1988).

[15]Kadambe, S., and Boudreaux-Bartels, G., "A comparison of wavelet functions for pitch detection of speech signals", Proc. of ICASSP, May 449-452, (1991).

[16]Shabana, W., " A wavelet-based pitch detector for musical signals " Technical report, Department of Mathematical Sciences, University of Bath, Oct., (1999).

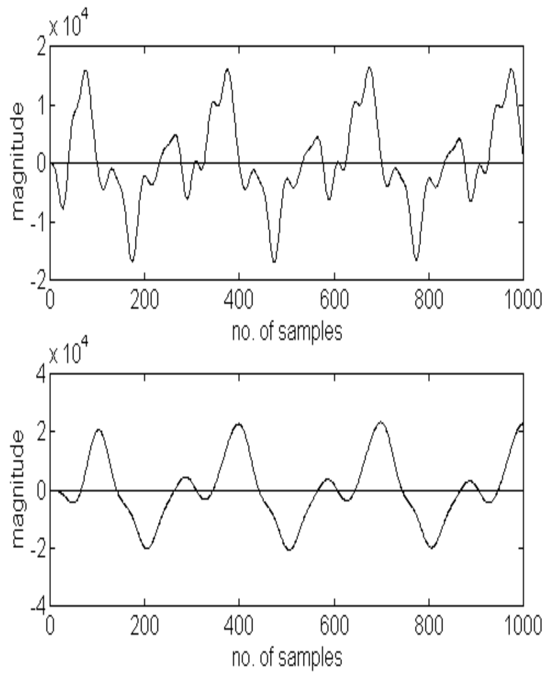

Figure 2. scale 2^5 and scale 2^6 of a of the guitar tone using quadratic spline wavelet

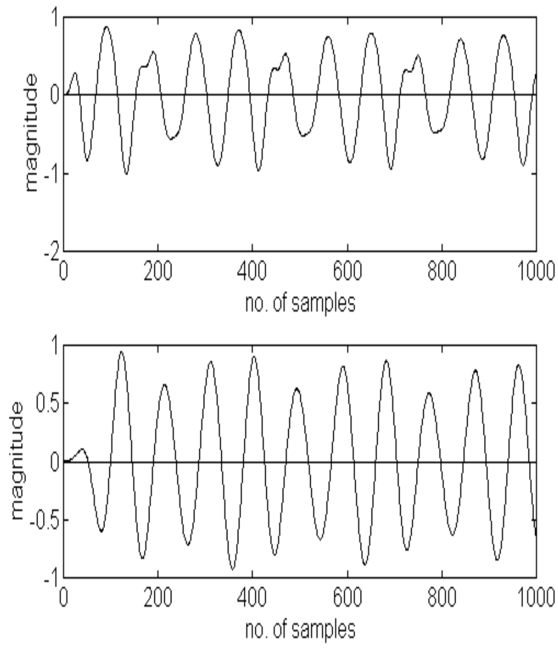

Figure 4. Scale 2^5 (top) and scale 2^6 (bottom) of a tanpura signal using quadratic spline wavelet

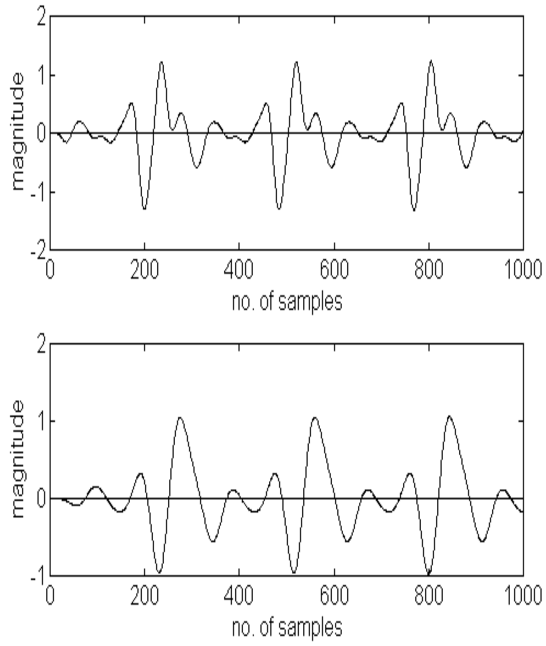

Figure 3. top: scale 2^5 and bottom: scale 2^6 of the sax signal using quadratic spline wavelet

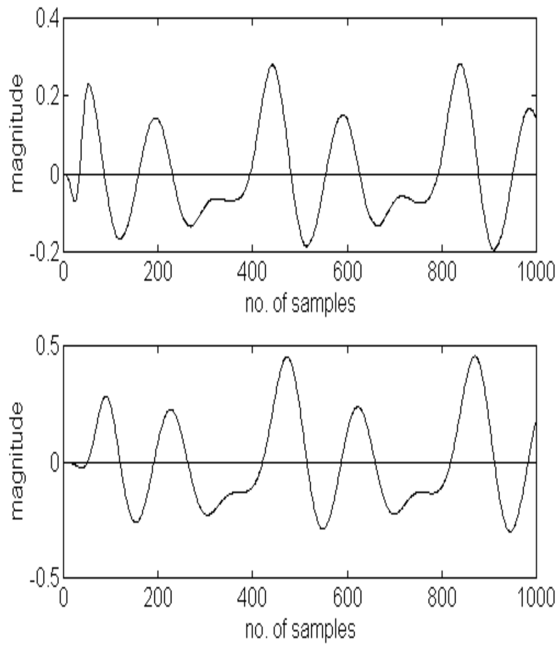

Figure 5. top: scale 2^5 and bottom: scale 2^6 of a singing voice signal using quadratic spline wavelet