VIRTUAL MUSIC TEACHER FOR MONOPHONIC MUSIC SIGNALS

Aberehe Niguse Gebru

Master of Industrial Science, Electronics Engineering, Faculty of Engineering Technology,
Campus Group T Leuven Andreas Vesaliusstraat 13, 3000 Leuven, Belgium

Supervisor: Koen Eneman
Faculty of Engineering Technology, Campus Group T, Leuven
Andreas Vesaliusstraat 13, 3000 Leuven, Belgium
Koen.Eneman@kuleuven.be

ABSTRACT

The target of this thesis paper is to develop digital signal processing algorithms and implement a software environment to detect the pitch of music signals. Analysis of music signals, detecting the pitch and tracking notes is the main part of the work. Different state of the art pitch detection algorithms were investigated and their advantages and disadvantages were studied and compared among each other. In this thesis, the Wavelet transform method is used for monophonic pitch detection. The proposed method was developed using MATLAB® and tested for various music tracks which were produced from multi-track MIDI and audio editing software and some downloaded from the net. MATLAB® graphical user interface is used to display the detected as a feedback to the musician.

Keywords
Autocorrelation, MATLAB®, Music education, Pitch Detection, Wavelet Transform

1   INTRODUCTION

Virtual Music Teacher mainly deals with detecting the fundamental frequency (f0) of a musical signal. The fundamental frequency is also called the pitch (perceptive frequency of sound) of the signal. The input musical signal is pre-processed and passed through different steps of the algorithm to increase the accuracy and efficiency of pitch detection.

The input signal may have a single fundamental frequency which can be represented by a pure sinusoid. Such signal is called a monophonic signal. The signal may also contain more than one fundamental frequency at the same time. Signals that contain more than one pitch at the same time are called polyphonic signals. The task of analysing the pitch content of polyphonic audio (polyphonic music, multi-talker speech, etc.) is known as multi-pitch analysis("Multi-pitch Analysis," n.d.). Multi-pitch analysis generally includes estimating frequency and number of pitches in each frame and organizing it to track into notes or continuous pitch segments. This paper addresses the different algorithms and methods used to detect the pitches of monophonic signals and discuss the proposed method in detail. The proposed method for monophonic signals relies on the wavelet transform.

1.1 Nature and scope of the thesis

Many people have the ability to hear and detect the pitches played by different musical instruments and even human voices. But there is no completely accurate and efficient automatic method for pitch detection yet. Monophonic signal pitch detection methods are more accurate and advanced compared to the methods/algorithms used for polyphonic pitch detection. Polyphonic pitch detection is still an unresolved area of research in which many methods have been implemented and tested but the results were not fully accurate. Accordingly there is no single accurate and efficient polyphonic pitch detection algorithm.

Virtual Music Teacher provides best pitch detection methods for monophonic signals. A graphical user interface (GUI) is implemented to display the detected pitches in the form of music notes. An apprentice musician will use this software environment to get feedback about the pitches played or about the notes within the music signal of interest inputted from file.

This will help learn the notes of music signals especially for those who are beginner music trainee and show their progress during their training. It will also give the music information of the signal which may also be used in automatic music information transcription systems.

1.2 Analysis of existing methods

As mentioned in the introduction part, many different pitch detection algorithms have been tested at different times in different places. These different methods differ in many aspects. Their computational efficiency, accuracy, delay and implementation costs are the main parameters used to measure these detection algorithms.

Generally these detection algorithms can be classified into three categories depending on the signals to which the algorithm is applied and relevance of the output. Some algorithms are applied on the original signal on which no further processing and transformation is applied. These algorithms are time domain algorithms. Mainly those algorithms apply convolution or correlation operations on the time signal. Some of the algorithms used in time domain are: Zero crossing, Autocorrelation, Maximum likelihood, etc("PITCH DETECTION METHODS," n.d.). Zero crossing method uses the number of times a signal crosses the zero level to estimate period and fundamental frequency. This method is simple which gives more reliable results for simpler signals. For complex signals which are harmonic with more partials the result is poor. Autocorrelation implements correlation operation to find the

similarity of a signal with a shifted version of itself. This method can be used to detect pitches of harmonic signals by first detecting the peaks of the result of the autocorrelation function. Harmonic signals contain peaks at the fundamental frequency and integer multiples of it. This method is an expensive method as it consists of many multiply-add operations. But this cost can be improved significantly using FFT method of calculating the autocorrelation of a signal. In this way of finding the autocorrelation result, the FFT operation is applied on the signal first and then the inverse FFT is calculated finally from which the autocorrelation peaks are going to be sought. The massive multiply-add operations are not part of this method of finding the autocorrelation function. This makes more computationally efficient method of autocorrelation manipulation. Maximum likelihood method can also be used in time domain signals. The signal is initially divided into many segments of some specified length. Peaks will be detected at specific points in each segment which will provide a constant distance/period between the peaks
.

The second category of methods contains algorithms which are applied on the frequency transform of the input signal. Before applying these methods the input signal is transformed into frequency domain using Fourier Transform or Wavelet Transform. The majority of these methods use per-frame operations. The input signal is broken into small frames, multiplied by suitable windowing function and finally the short time Fourier transform of the frame is calculated. The peaks of the frame will be separated and will be selected for further operation to estimate the most probable fundamental frequency of the signal. Some of the algorithms which are included in this group are: FFT, Harmonic frequency spectrum, Cepstrum, Wavelet transform, etc.

FFT method can be used to effectively detect pitch of signals mainly which are not harmonic and complex in nature. The first maximum from the spectrum will be taken and the corresponding frequency is taken as fundamental frequency. This method is more effective for signals with highest amplitude peak found first in the spectrum of the FFT as shown in Figure 1, taken from("Pitch Recognition with Wavelets," n.d.). The main drawback of this method is when peaks of the higher partials have higher amplitude compared to the fundamental peaks. Many harmonic signals may have higher amplitude harmonic components, in which applying FFT method will give the wrong pitch. FFT is widely used technique but it has limited time and frequency resolution which is subjected to uncertainty principle("Uncertainty Principle," n.d.). In this method the bin number with the highest magnitude will be used to calculate the fundamental frequency according the following equation:

$$\left(-\right)$$

Where fs is the sampling frequency, N is the length of the segment or frame and n is the bin number corresponding to the highest amplitude value.The drawbacks of FFT are improved using the Harmonic Product Spectrum method (HPS)("Efficient Pitch Detection Techniques for Interactive Music," n.d.). HPS can be used to detect accurately even if the higher partials of a signal have higher amplitudes than the fundamental frequency as shown in Figure 2("Pitch Recognition with Wavelets," n.d.). The frequencies of all the harmonics are calculated and the highest common divisor of the calculated values will be taken as the fundamental frequency.
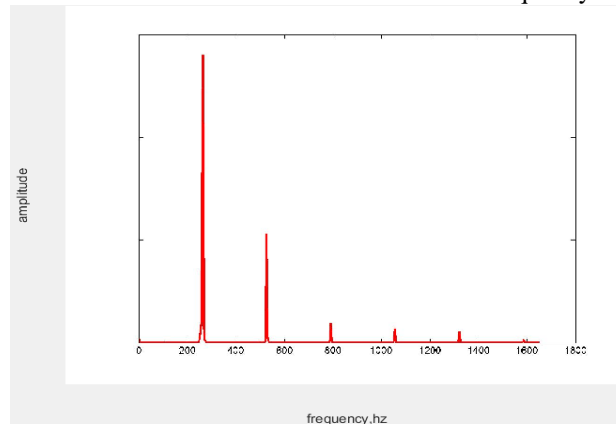

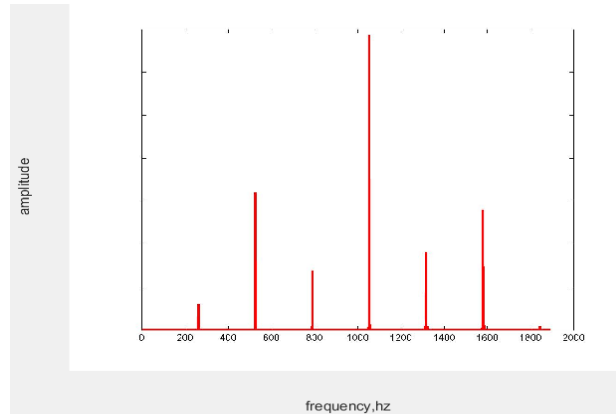**Figure 1: fundamental with highest magnitude**

Figure 2: fourth harmonic with highest magnitude

Cepstrum is another frequency domain pitch detection mechanism which applies Fourier Transform to the log-magnitude Fourier Transform of the signal. If the signal is harmonic, it will show periodicity in the frequency representation, thus taking FFT again will show a peak corresponding to the period in frequency of the signal.

Wavelet method is the proposed monophonic pitch detection method implemented in this thesis. The main drawback in FFT and HPS methods is that there is constant window length. There is less accuracy and limited resolution due to the fixed window length. But wavelet method implements logarithmic nature of music, which provides multi-resolution analysis of the signal at different scales. This multi-resolution property of wavelets gives improved resolution at some frequency ranges and improved computation time. Due to this advantage Wavelet detection method is applied for the monophonic part. Besides, the continuous wavelet transform is chosen because of its fine grained resolution and redundancy analysis capability. Discrete wavelets are used in fast and non-redundant transforms with good space saving. But, the Continuous wavelets have the ability to detect small discontinuities and frequency variation of a time series signal. The details of the process will be explained in a later section.

The third and last group includes methods used from both time domain and frequency domain approaches which are also called spectral/temporal methods. These methods include processes from both domains which helps correct errors of one domain by processing in the other domain.

.

1.3 Organization of the paper

In the introduction, the problem statement of the thesis is highlighted. Besides, the overview of the different state of the art pitch detection algorithms, their classification and comparison is explained for both monophonic and polyphonic music signals.

In the materials and methods section, the full detail of the proposed detection algorithm for monophonic pitch signals is presented and then each of the steps and processes will be explained.

In the results section, the results of the various state of the art detection methods is compared with the results obtained in the proposed methods. In the discussion section, a generalized discussion is presented on the proposed detection methods, the advantages and drawbacks of the methods, the theoretical implications and practical application of the approaches and results is paraphrased and addressed. The downside of the methods is explained and directions are forwarded for future works and improvements for researchers who are interested in the field. Finally, the paper will conclude with conclusion, containing the summary of the work in this thesis.

2 METHODS

The wavelet algorithm for monophonic pitch detection is explained in detail in this section.

2.1 Wavelets

As already mentioned in the introduction part, due to the discussed reasons the continuous wavelet transform(Alfred Mertins, 1999) is used for detecting the pitch of monophonic music signals. The continuous Wavelet Transform analysis redundancy behaviour tends to reinforce traits and makes information of the signal more visible("Continuous and Discrete Wavelet Analysis," n.d.). The continuous wavelet transform is defined by the following formula:

$$\int f \quad \frac{}{\sqrt{a}} \quad \left(\frac{}{}\right) dt$$

Where:

    f(t) = signal
        = Mother wavelet
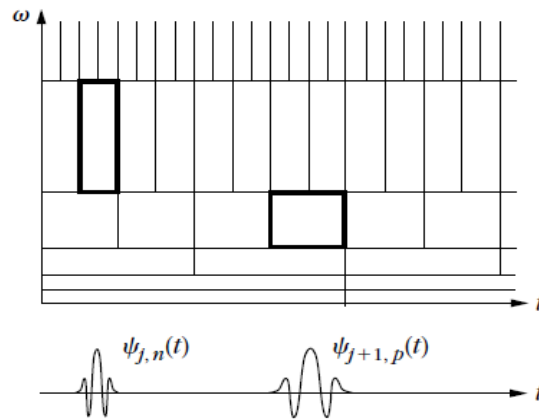   a = scaling factor
    b = shift parameter

Figure 3: wavelet resolution

Figure 3 shows the time and frequency resolution property of wavelet analysis. This is the main advantage in time and frequency localization of wavelet transform processing. This will be discussed in the discussion section of the paper.

One of the most simple and suitable wavelets is the Haar wavelet. The Haar wavelet is an orthogonal wavelet in which the product between corresponding wavelets gives zero. Haar wavelet mother function/mother wavelet is represented by the following Mathematical formula (equation 3). The value of the mother wavelet is unity on a small horizontal scale.

$$\left\{-\right.$$

The values and the exact locations in time depend on the scaling and shifting parameters ('a' and 'b' respectively, see equation 2).

The Haar wavelet Transform scaling function and mother wavelet have the following appearances (figure 4
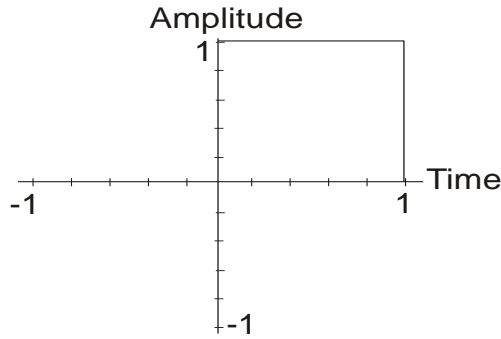
and Figure respectively).

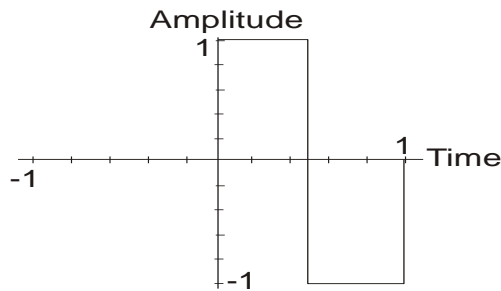**Figure 4: Haar wavelet scaling function**

Figure 5: Haar mother wavelet

When the scaling parameter a, in the wavelet transform is varied it has an effect of stretching and compressing the mother wavelet. When the wavelet transform is performed using the Haar system on the input signal, the signal is converted in to a series of wavelets corresponding to the values of the scaling and shift parameters. In this way the wavelet transform provides a signal analysis with frequency and time location is bounded. The operated signal is stored efficiently with more localized time and frequency. The signals are better approximated and accurately localized for better further processing. The wavelet transform block is shown in the second block of the general block diagram representation of the wavelet monophonic pitch detection algorithm in Figure 6.

The wavelet Transform function uses the wavelet toolbox from MATLAB®. The output of the wavelet process has many advantages for the later processes. The periodicity and discontinuities of the signal become more visible and detectable. The effects of noise and low energy portion of the input signal are suppressed and portions with higher amplitudes become into a clear picture. The wavelet process is performed in different scales. The first scale is almost the same as the original input time signal. As the scale of the wavelet increases, the processed signal at that specific scale will have clear peaks and discontinuities which help identify the periodicity and peak values of the signal. This property of Continuous wavelet transforms is especially important to detect abrupt changes and frequency variations of signals.
.
The third block in the block diagram (see Figure 6) performs a peak detection operation on the signal
received from the wavelet processing block. The signal has suitable characteristics to simply and clearly identify the prominent peaks. Using correlation of the signal obtained from wavelet block before finding the peaks provides best and accurate results.

The variation of the signal with the wavelet analysis scale value is depicted with the following figure (figure 6). The figure shows that the original signal analysed (subplot of first row) is almost similar to the continuous wavelet analysis output at the first scale (subplot of the last row). As the scale increases the signal becomes more readable and the discontinuities are more visible as this property of the wavelet was explained in the introduction section. The second row from the top indicates the highest scale in the given analysis.
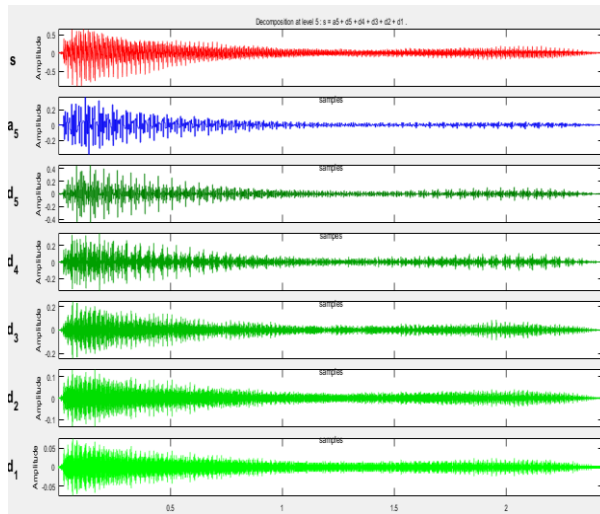
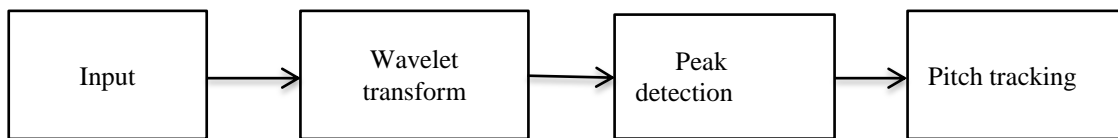Figure 6: the input signal and its waveforms at different wavelet scales.

Figure 7: block diagram of wavelet pitch detection

Peak values which are above some specific threshold are taken in to comparison and the ones that are local maxima are taken as peak values. The vector signal that contains the detected peaks is used as input to the pitch tracking block i.e. the last block in figure 7.

The last step in the detection process is detecting the pitch of the signal as shown by the pitch tracking block in the wavelet block diagram. Considering the periodicity detection property of the wavelet and the obtained peak values from the peak detection algorithm, the period of peaks is obtained. The period is used to calculate the fundamental frequency of the signal. The period value that exists frequently in the signal is used to define the frequency feature of the signal. The inverse of the detected frequently existing period is taken as the fundamental frequency/pitch of the signal. Finally this detected pitch is converted into note names corresponding to the names in equal tempered scale piano.

3   RESULTS

3.1 Monophonic

As it is discussed in the method part of the paper the monophonic pitch detection analysis is done using the wavelet transform method. But, in addition to the wavelet analysis it was also tested using some other methods like FFT method, Autocorrelation method and zero crossing methods. The output of the result is displayed using MATLAB® graphical user interface as shown in figure 10. In the MATLAB® graphical user interface the load button is used to select the music track from the folder that contains the current workspace. Once a track is loaded it is analysed by selecting one of the methods shown in the interface. The output feedback displays the note name and corresponding midi number of the note. There are four detection methods represented i.e. monophonic (wavelet), FFT, auto (autocorrelation) and polyphonic (summary autocorrelation). The wavelet algorithm is tested for different notes of

different instruments produced using anvil midi studio("Anvil Studio," n.d.). iTunes was used to convert midi files into wav formats. Notes from the whole note range are tested. It almost always detects the correct pitch for signals with pure note with less harmonics and noise. Many tones are tested starting from the lowest to notes in the high range and the results are correct except some single semitone deviation for the lowest octaves. The note which is immediately one semitone above it is detected for some of the notes in the first octave (e.g. C#1 instead of C1).

FFT methods are also tested for the signals and gave good results for signals which have the highest amplitude peak at the first peak/fundamental frequency like in vocals. But generally FFT is didn't give good detection results. For signals that have higher harmonics with higher magnitude compared to the first harmonic the results are not the correct ones. It detects one of the higher harmonics. This drawback of FFT can be improved using harmonic product spectrum method which is not tested in this thesis.

Autocorrelation method is also tested giving good overall detection results. This method gives accurate detection results even for the notes of the lowest octaves. But it doesn't give good detection for complex signals and it has not relevant computational significance. Similarly zero crossing method was tested for some notes giving accurate results a few for pure tones and simple signals but generally its detection is poor.

Additionally the wavelet method was tested for notes of other musical instruments like violin, cello, viola, bass guitar and guitar("Note Frequencies," n.d.) and("Anvil Studio," n.d.). The method perfectly detected the pitches of the note samples taken from each of the musical instruments.

Method	F0-detection
FFT	60%
Autocorrelation	87%
Wavelet	90%

Table 1: detection rates of the three methods according to the samples tested (the result is obtained from 80 test samples of Electric grand produced by midi software("Anvil Studio," n.d.)).

Zero crossing method only gives good results for some specific samples tested. But generally the result didn't give good results.
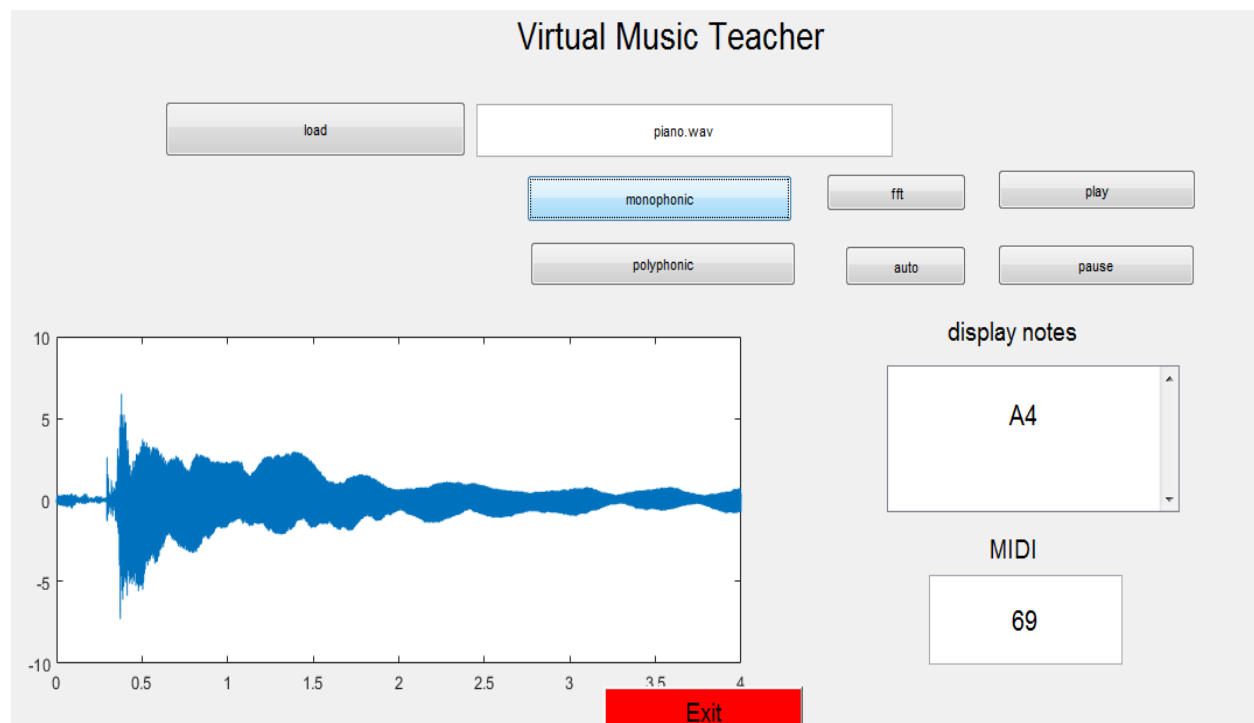

Figure 10: screenshot of the matlab gui implemented

4 DISCUSSION

The monophonic pitch detection is the main part of this paper as described in detail in the method section of the paper. The graphical user interface is the last part which is used to display detected pitches as final feedback to the music learner.

Monophonic pitch detection is an almost solved problem with many methods implemented and showed relevant results. For simple/not complex music signals, time domain pitch detection mechanisms give good detecting capabilities with efficient computational cost. But for signals with more complex harmonics and noisy situations the results are not good enough. Signals which are more complex are efficiently analysed with the more widely used frequency domain approaches that depend on FFT application. By FFT, it most of the time, means using the STFT analysis. STFT operates on frame by frame bases of the original whole signal. The main downside of STFT is that the window length of the frame is fixed. This fixed window length gives a fixed resolution in the frequency domain which gives constant absolute error. This error may not be acceptable for some frequency values. For example, 12 Hz absolute error from 1000 Hz may be tolerable, but 12 Hz error from 80 Hz is no more tolerable.

The wavelet analysis has better properties in some aspects compared to the Short Time Fourier Transform. The salient advantage of wavelet transform is their adaptive resolution. They have a constant relative error in the frequency domain, not constant absolute error unlike STFT. This makes the wavelet method more accurate in detecting the right pitches. The wavelet toolbox is applied for the work that uses wavelet analysis. This means the wavelet implementation used classical wavelet transform which is not computationally efficient compared to second generation wavelets (fast lifting wavelet Transforms)(Ergun Ercelebi, 2002). Generally the wavelet method gives good detection results with comparable computational efficiency to the other methods tested.

5 RECOMMENDATION

The wavelet method used in monophonic detection is classical wavelet transform from wavelet toolbox as explained in the discussion section. This classical wavelet analysis method has massive computational operations which makes it un-efficient in processing delay and computational costs. This may not be ideal for speed dependent applications like real time implementations. For fast real time applications fast lifting wavelet transforms are better choice. These transforms perform simple addition-subtraction operations on the coefficients in time domain. There is no time consuming complex operation in this type of wavelet mechanism.

One other thing which may also be of point of interest for future researchers in this work is implementing the algorithm in real time systems. Currently, there is massive computation and processing which doesn't make it suitable for real time implementation. This can be optimized and implemented in future works.

A larger window size gives better frequency resolution and decreases errors and a small window size gives better time resolution. But a larger window size means a large computation within a single window. This increases the computational cost per frame and may increase processing delay. There may be a way to deal with other methods to solve this time and frequency resolution offset against window length.

6 CONCLUSION

The VMT application integrates detection of the pitches of musical note and a user interface environment to display the detected pitches as feedback to a user/music learner. The GUI is developed using MATLAB® graphical user interface development environment (GUIDE).

Monophonic pitch detection is the heart of this work. The method used for monophonic detection is wavelet transform analysis. This method gives accurate results with better resolution. In addition to wavelet other methods are tested such as, FFT, autocorrelation and zero crossing methods. The wavelet analysis doesn't have computational advantage over the other methods, because the classical wavelet is used which includes massive computational operations. This can be further improved using second generation wavelets or even other new wavelets can be developed as it is mentioned in the recommendation section.

7 ACKNOWLEDGEMENTS

8   REFERENCES

Adrian von dem Knesebeck, Sebastian Kraft and Udo Zolzer. (2011). Real time system for backing vocal harmonization. Retrieved April 29, 2016, from http://recherche.ircam.fr/pub/dafx11/Papers/52_e.pdf

Alfred Mertins. (1999). Wavelet Transform. Retrieved April 29, 2016, from http://dsp-book.narod.ru/493.pdf

Anvil Studio. (n.d.). Retrieved May 6, 2016, from http://www.anvilstudio.com/

Continuous and Discrete Wavelet Analysis. (n.d.). Retrieved May 4, 2016, from http://nl.mathworks.com/help/wavelet/examples/continuous-and-discrete-wavelet-analysis.html

Efficient Pitch Detection Techniques for Interactive Music. (n.d.). Retrieved May 9, 2016, from https://ccrma.stanford.edu/~pdelac/research/MyPublishedPapers/icmc_2001-pitch_best.pdf

Ergun Ercelebi. (2002). Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals. Retrieved April 29, 2016, from http://ac.els-cdn.com/S0003682X02000555/1-s2.0-S0003682X02000555-main.pdf?_tid=d46220e4-0e00-11e6-be85-00000aacb35e&acdnat=1461930920_a07f8cd1803b0831eb2f54ac981bc804

k-means clustering. (2016, May 9). In *Wikipedia, the free encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=K-means_clustering&oldid=719385961

Matti Karjalainen. (2015). Auditory Interpretation and Application of Warped Linear Prediction. Retrieved April 29, 2016, from http://www.ee.columbia.edu/~dpwe/workshops/crac/papers/karjalainen.pdf

Matti Karjalainen,ero Tolonen. (n.d.). Multi-Pitch And Periodicity Analysis Model. Retrieved May 9, 2016, from http://www.ee.columbia.edu/~dpwe/papers/KarjT99-pitch.pdf

Multi-pitch Analysis. (n.d.). Retrieved May 9, 2016, from http://www.ece.rochester.edu/~zduan/multipitch/multipitch.html

Note Frequencies. (n.d.). Retrieved May 6, 2016, from http://www.seventhstring.com/resources/notefrequencies.html

PITCH DETECTION METHODS. (n.d.). Retrieved May 9, 2016, from https://ccrma.stanford.edu/~pdelac/154/m154paper.htm

Pitch Recognition with Wavelets. (n.d.). Retrieved May 9, 2016, from http://studylib.net/doc/10118812/pitch-recognition-with-wavelets

Ray Meddis and Lowel O'Mard. (1997). A unitary model of pitch perception - Meddis_OMard 1997(Pitch).pdf. Retrieved April 29, 2016, from http://essexpsychology.macmate.me/HearingLab/publications_files/Meddis_OMard%201997(Pitch).pdf

R. Meddis and M. J. Hewitt. (1991). Virtual Pitch and Phase Sensitivity of a Computer Model of the Auditory Periphery. Retrieved April 29, 2016, from http://www.essex.ac.uk/psychology/department/hearinglab/publications_files/Meddis%20%26%20Hewitt%201991%20pitch%20I.pdf

Sebastian Kraft, Udo Zölzer. (2014). Polyphonic pitch detection by iterative analysis of the autocorrelation function. Retrieved April 29, 2016, from http://www.dafx14.fau.de/papers/dafx14_sebastian_kraft_polyphonic_pitch_detectio.pdf

Uncertainty Principle. (n.d.). Retrieved May 9, 2016, from https://www.dsprelated.com/freebooks/mdft/Uncertainty_Principle.html

9   APPENDICES

The appendices of the paper includes a CD with a folder that contains many MATLAB® functions implemented for monophonic pitch detection, gui MATLAB® script which was developed using MATLAB® GUIDE and many .wav files which were used to test the developed algorithm.