Design and Implementation of Real Time Audio Pitch Shifting on FPGA

Pranita R Morbale[1], Mahesh Navale[2]
[1]PG Student, SKN College of Engg,Pune , [2]Assistant Prof. SKNCOE, Pune, Research Scholar Karpagam University, Coimbatore,

Abstract

Pitch shifting is a process to transpose tone up or down without changing its periodic properties. Current multimedia field is developing very fast where sound manipulation is important task. Mainly there are two categories to develop pitch shifting algorithm first one is time domain and another one is frequency domain. Time domain algorithms are not efficient for polyphonic audio signals .Frequency domain algorithm provides a better time-frequency resolution. In proposed system two algorithms are compared STFT (Short Time Fourier Transform) and CQT (Constant Q Transform).Both are analyzed in frequency domain. The performance of both algorithms is evaluated using MATLAB Simulink platform. It shows CQT is better time-frequency resolution, It is selected to deploy on FPGA evaluation board. Now-a-days FPGA (Field Programmable Gate Array) technology has become a viable target for the implementation of real time algorithms in different applications. Main objective is to design an algorithm efficient enough to enable real-time operation on FPGA board that allows real-time microphone input and speaker output. With Audio D/A and A/D conversion performed using AC'97 codec which is on development board. Transform (STFT) which is developed in MATLAB for Simulation purpose. As Field Programmable Gate Array (FPGA) technology has become viable target for real time algorithms implementations in different applications. Hence CQT is developed in MATLAB SIMULINK environment for hardware co-simulation and HDL is downloaded in Vertex 5 FPGA kit for real time implementation

Keywords

Pitch, STFT,FPGA, MATLAB Simulink,AC'97 codec, Virtex 5

1. Introduction

Shifting of pitch is a audio processing that changes the pitch of a sound or music without changing its temporal component. That is, scaling factor is constant which Q is irrespective of frequency of signal. .It can be defined as ratio of center frequency to bandwidth. Pitch-shifting main two stages which involves 'Time scaling' and after that to get original signal 'resampling' .Time-scaling can either be performed in the time-domain (TD) or the frequency domain (FD). Approaches operating in the frequency domain are often based on the phase vocoder where time-scaling is achieved by altering adding or deleting frames .Phase vocoder operates in the frequency domain while PSOLA operates in the time-domain [2][3]. The Time-domain algorithms attempt to determine the pitch directly from the speech waveform while frequency domain algorithms use some forms of spectral analysis to determine the pitch period. Pitch changes, pitch scaling, or pitch modification means transposing the pitch without changing the characteristics of the sound. In addition, it is defined as the process of changing the pitch without affecting duration. Phase vocoder-based pitch shifting implementations usually operate on the Short Time Fourier Transform (STFT) representation. Rigid time-frequency resolution is main disadvantage of STFT [8].Which is not suitable for polyphonic and dense music signal. Where flexibility in resolution is required. Ideally frequency resolution should increase from low to high and time resolution should increase form high to low. Using the STFT, two neighboring peaks can not be distinguish from each other results in excitement of only one spectral peak. The time-resolution at higher frequency regions might be not fine to track quick temporal changes.

In the DSP age, speed and pitch can be altered independently. One can increase speed without changing pitch as in tone without changing its time component [1] If system has stationary and uniform signal then it's trivial. For dense polyphonic music signals, computational complexity increases as frequency translation is required for every spectral peak. This is because the STFT frequency bins are linearly spaced whereas scaling all frequencies by a constant factor corresponds to a constant shift on log-frequency scale. Thus there are number of algorithms to date. Past work is done on

Fourier Transform but due to its linear spacing bins it's not preferable for processing of speech and music signals. Hence invertible constant-Q transforms (CQT) is proposed which has high Q-factors with geometrical bin spacing [1].

2. System Model

In this system, pitch shifting is described using wavelet based technique called as constant Q Transform (CQT). The purpose of this design project is to implement a real-time speech pitch shifter on an FPGA Pitch shifting is altering speech to change the pitch without changing its duration. Frequency domain pitch shift delivers high quality and low noise Audio output with keeping time component same and without changing formant.

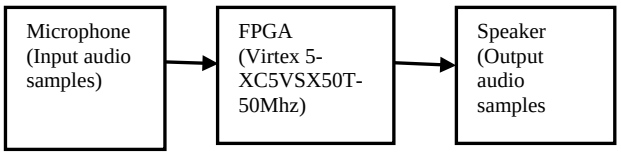


Fig.1. System Overview

Figure 1 shows pitch shifting system laid on the FPGA which is elaborated in [5].

3. Previous Work

In FD pitch shifting frequency translation is done with constant scaling factor. To understand the efficiency of invertible CQT here various methods are discussed to get each aspect of pitch shifting..

A. Background

The classical approach to pitch shifting is to first compute the STFT time-frequency representation of a signal. Afterwards, the grid spacing and coefficient phases are scaled to create a synthesis grid. The inverse is then computed to reconstruct the signal. It is summarized as follows:[1]

- *Compute the STFT representation of the signal.*
- *Convert coefficients into polar form*
- *Unwrap the phase and divide by the scaling factor C.*
- *Reconstruct signal using new synthesis scale. Time*

B. Pitch shifting methods

In many applications sound synthesis is important. There are two main methods to pitch shift.

1. *Time Domain Pitch Shifting (TDPS):* There are no of TD (Time Domain) algorithms which involves simply frequency translation and resampling. These does produce true pitch shifts but duration get changed which is not desirable [8]. The effect of this framing on the signal can be elaborated after taking Fourier Transform (FT) of signal which will show harmonic distortion of signal [1].

2. *Frequency Domain Pitch Shifting (FDPS):* Same as the time-domain approach, the frequency-domain approach is based on shifting small overlapping windowed blocks of data in time and resampling. The overlapping data blocks are shifted closer together to create a compact signal with the same pitch but shorter duration. To resample that signal it should be expand to its original duration, this stretching of duration will shift down the pitch. The resulting phase inconsistency introduce jitter, same as time-domain technique. In frequency domain phase information can be updated to eliminate inconsistency.

C. Pitch shifting Algorithms

There are number of algorithms which are used to shift pitch which use both FD (frequency Domain and TD(Time Domain) methods to manipulate signal .

a) Pitch Synchronous Overlap and Add (PSOLA)[3]
b) Phase Vocoder [5]
c) *Short Time Fourier Transform (STFT)* [5]
d) Wavelet transform [3].
e) *Constant Q Transform (CQT)* [1][8]

4. Proposed Methodology

Among STFT, CWT and phase Vocoder techniques STFT is efficient and easy to implement. Hence frequency domain STFT is implemented in this design. And due to high quality audio signal CQT is also implemented here.

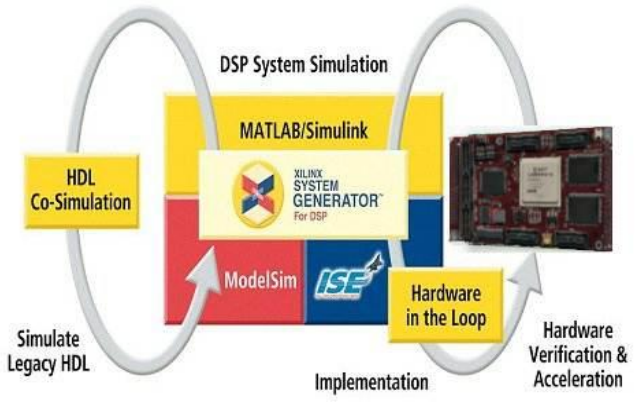


Fig.2 Complete System Development Process

Design flow involves software developments in MATLAB Simulink environment and its deployment on hardware of FPGA board. Fig 2 shows integration of both software models and hardware platform.

A. Software *Implementation*

In this paper, pitch shifting using STFT and CQT are simulated in MATLAB Simulink platform to analyse its performance metrics .After analysing efficient one is deployed onto FPGA evaluation word with hardware co-simulation.

1. FD STFT*:* Mathematically, the STFT looks like this the Fourier transform of the windowed signal x[n]W[n-m]:

$$X[k] = \sum_{n=0}^{N-1} x[n]W[n]e^{-j\left(\frac{2\pi kn}{N}\right)} \quad (1)$$

Shifting window in temporal component will increase value of m. When dealing with music in real-time assume it is employed with the first N samples at any instant of time; so the window starts and stops with the samples at n=0 and

n=N-1

$$X[m,\omega] = \sum_{-\infty}^{+\infty} x[n]W[n-m]e^{-j\omega n} \quad (2)$$

a) *Process:* To avoid phase discontinuity in phase vocoder algorithm, FD STFT is developed to do phase update and filter out smearing effect. Figure 3 explains stepwise description in this algorithm development.

b) Diagram

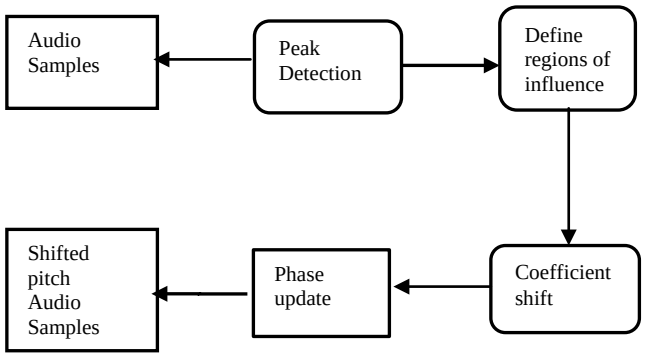


Fig. 3.FD STFT Process

c) *Drawbacks:* FD STFT can update phase to avoid inconsistency of phase, but its output is not perfectly in phase. It has some cons as follows.

1) The resolution of the windowing function W[] is fixed for all frequency values. Which does not give flexibility.

The standard STFT has equally spaced frequencies because the exponent increases linearly with k.

2) It gives rigid response of time and frequency resolution. Both these are terribly suboptimal because the ear's response is logarithmic, not linear.

2. FD CQT: It is assumed that all phase values depending on the peak's phase value forms influence region are dependent on peak's horizontal and vertical phase can be retained exactly for a constant-frequency sinusoid. Linear spacing bin problem is resolved in FD CQT by developing geometrical spacing bins. The pitch-shifting algorithm proposed here is based on the CQT implementation proposed in [1] [8]. *Process:* Figure 4 explains FD CQT algorithm development process. CQT-based pitch-shifting that transforms a time-domain signal x(n) into the time-frequency domain so that the center frequencies of the frequency bins are geometrically spaced and their Q-factors are all equal.

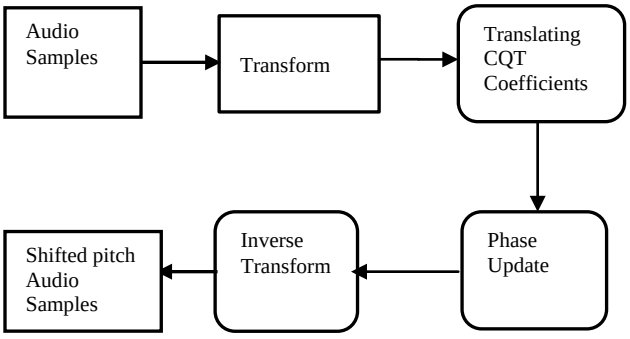


Fig. 4. FD CQT Process

a) *Signal Model of CQT:* The CQT transform $X^{CQ}(k,n)$ of a discrete time-domain signal x(n) is defined by

$$X^{CQ}(k,n) = \sum_{j=n-\lfloor N_k/2 \rfloor}^{n-\lfloor N_k/2 \rfloor} x(j)a_k(j-n+N_k/2) \quad (3)$$

Where k = 1; 2,……,K indexes the frequency bins

of the CQT,

b_c denotes rounding towards negative infinity

$a_k(n)$ Denotes the complex conjugate of $a_k(n)$.

The basic functions $a_k(n)$ are complex-valued waveforms, here also called time-frequency atoms and are defined by

$$a_k(n) = \frac{1}{C}\,\omega\left(\frac{n}{N_k}\right)\exp[i(2\pi n\,\frac{f_k}{f_s} + \emptyset_k)] \quad (4)$$

Where f_k is the center frequency of bin k

fs denotes the sampling rate

w(t) is a continuous window function

N_k isThe window m function is zero outside the range

$$N_k = \frac{fs}{fk\,(2\pi - 1)} \quad (5)$$

Φ_k is a phase offset and _k = 0 for the transform

proposed in [8].

C is a scaling factor

$$C = \sum_{l=-\lfloor \frac{N_k}{2} \rfloor}^{\lfloor \frac{N_k}{2} \rfloor} \omega \quad (6)$$

b) *Advantages:* CQT proposes the logarithmic frequency bin spacing of the CQT which is preferable in audio signals as ear has logarithmic response. There are many advantages over STFT as follows.

1) Due to the logarithmic frequency resolution of the CQT, relative detuning between partials is avoided and closely spaced sinusoidal components at lower frequencies can be distinguished.

I2) It reduces phasiness of time scaling and artefacts in the output signal due to high resolution.

The software which are used to design this system are listed below.

- MATLAB 13a
- Xilinx ISE 14.1,Xilinx 11.1
- HTTP

B) MATLAB Simulink models:

1) Pitch Shifter based on STFT

The Short-Time Fourier Transform (STFT), a common audio processing tool that involves taking the Discrete Fourier Transform (DFT) of short, periodic blocks of an audio signal. Figure 5 and figure 6 shows the models of STFT based down and up shifter respectively.

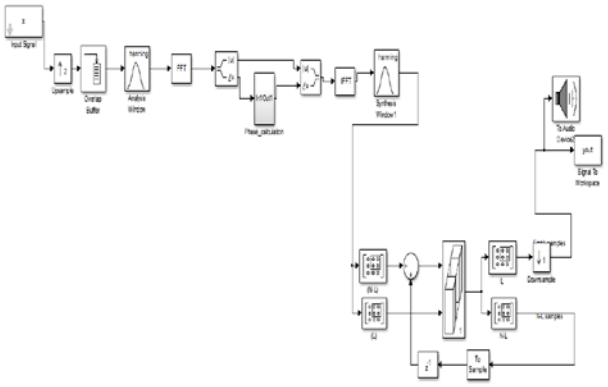


Fig. 5 Simulink Model of STFT based DOWN Pitch Shifter

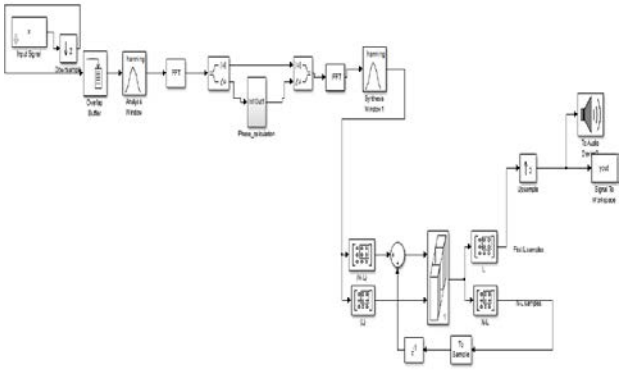


Fig. 6 Simulink Model of STFT based UP Pitch Shifter

2) Pitch Shifter based on CQT

Due to incoherency in phase in output audio in STFT pitch shifting technique, it gives transients and smearing effects. Hence CQT is implemented which gives logarithmic frequency resolution instead of rigid time-frequency resolution. After analyzing the two pitch shifting algorithms, CQT approach was selected reasonably high quality output and low artefacts. CQT down and up shifters are shown below figures namely figure 7 and figure 8.

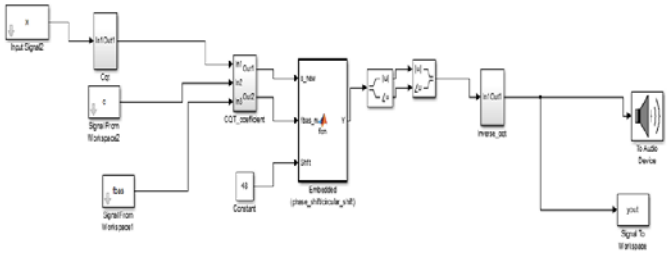


Fig. 7 Simulink Model of CQT based Down Pitch Shifter

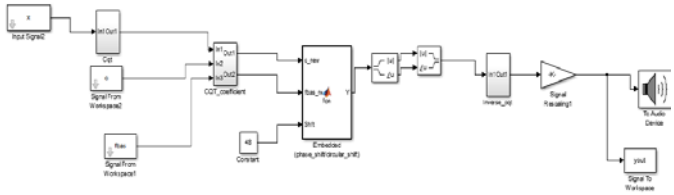


Fig. 8 Simulink Model of CQT based UP Pitch Shifter

C*)* Hardware Implementation:

This design is implemented on FPGA due to current growth of VLSI technology. Two algorithms are developed for analysis purpose. Its hardware Requirements are listed below.

- Virtex-5 FPGA board
- Xilinx USB downloads cable.
- Serial to parallel cable
- Mike- having 20 db gain

1) Modeling process

Moving from an algorithmic description to a quality, cost effective FPGA solution is anything but trivial. The CQT algorithm was originally developed as MATLAB scripts using high precision, complex floating-point arithmetic. MATLAB 13a with the fixed-point tool box were used in this process. The fixed-point implementation was then tested against the floating-point version. Following validation of the fixed-point implementation, a Xilinx System GeneratorTM (XSG) model was developed. Xilinx System GeneratorTM is an FPGA hardware DSP development environment that sits above MATLAB and Simulink software packages [4]. The XSG package contains predefined blocks that can be readily compiled into a hardware description language (HDL) and subsequently synthesized for specific Xilinx FPGAs. An XSG model is certainly the fastest way to implement the complex functions of the algorithm, such as the fast Fourier transform (FFT).The verified hardware design was then synthesized, using XilinxISE 11.1 tools, and implemented on a high-end Xilinx Virtex-5 FPGA. For hardware development work the ML 506 development board (containing a Virtex-5 device) was used. Figure 10 shows whole process of XSG model generator from Simulink model . Appendix I is complete XSG model.

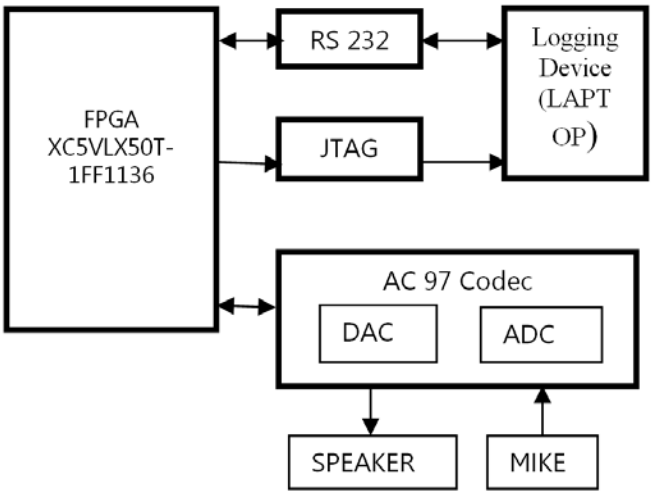


Fig. 9 Experimental set up

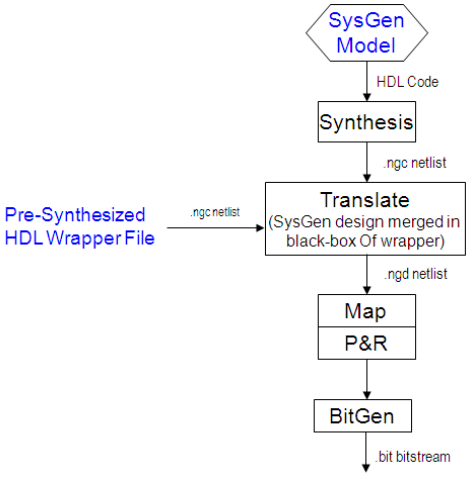


Fig. 10 Flow of .bit file Generation

2. Simulation/Experimental Results

MATLAB Simulink models of both STFT and CQT are simulated for up/down audio shifting. To decide which algorithm is more efficient, three performance metric parameters are calculated from output of each algorithm. The three parameters are namely PSNR, MSE and Max Error are defined in the Simulink model shown in figure 11. After getting pitch shifted output ,MATLAB script to calculate these parameters is executed, which takes original audio and pitch shifted output as input to the model.

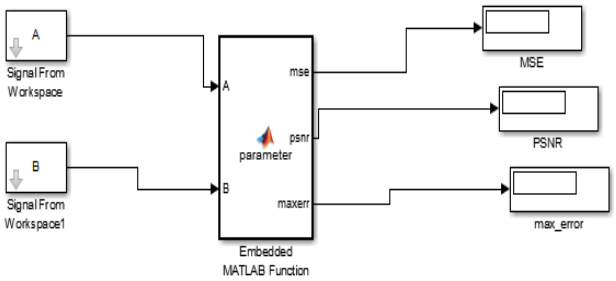


Fig.11 MATLAB Simulink Model for Parameter Evaluation

For verification purpose various audio file have taken. Some of them are real time and some are saved .wav files. Table 4.1 shows the parameter listing for various audio samples.

Table-1: Performance Metric Parameters of Output Audio Signals in MATLAB Simulink

Audio Samples	CQT			STFT		
	PSNR	MSE	Max Error	PSNR	MSE	Max Error
as1.6	69.72	0.000231 3	0.1724	67.19	0.0421	0.2796
As2.6	87.5	0.000210 2	0.1005	78.68	0.0423	0.1921
Sp.10	72.32	0.000100 1	0.1123	70.19	0.0621	0.2123
handel	76.11	0.000192 8	0.1231	72.08	0.0792	0.2123
Word_real	72.29	0.000143 8	0.1220	66.97	0.0864	0.2733

By analyzing table 1, it can be stated that CQT gives perfect reconstruction of input audio samples. It introduces phase update due to which it gives high PSNR compared to STFT which has rigid frequency resolution. Hence at this point CQT is decided to implement on the FPGA.

Figure 11 shows the final experimental set up for implementation of CQT on Virtex-5 FPGA.

After successful experimentation of CQT algorithm in MATLAB Simulink now it is ready to implement it on FPGA Virtex-5-ML506 evaluation board for real time implementation on hardware.

a) Synthesis Results

Synthesis is a process by which an abstract form of designed circuit behavior or register transfer level (RTL) has been converted into a design implementation i.e., in terms of logic gates. The synthesis of VHDL code has been carried out by Xilinx Synthesis Technology (XST) tool, which is part of Xilinx ISE 11.1 software.

Fig.12 Experimental set up

b) RTL schematic

This is a schematic representation of the pre-optimized design shown at the Register Transfer Level (RTL). This representation is in terms of generic symbols, such as adders, multipliers, counters, AND gates, and OR gates, and is generated after the HDL synthesis phase of the synthesis process.

c) FPGA Resource Utilization Summary

Table 4.2 shows a summary of the key FPGA resources used in this implementation. The following paragraph explains the terms used. DSP48 slices are dedicated specialized hardware arithmetic blocks specifically tailored for the efficient implementation of complex mathematical and DSP functions. Digital Clock Manager (DCM) blocks are used to manage

clock generation, distribution and to minimize clock skew. Block RAM (BRAM) are flexible blocks of RAM embedded in the FPGA fabric that can be utilized in a wide variety of ways, e.g. dual port or FIFO buffers, shift registers, large look-up tables (LUTs) etc. Slices represent the basic FPGA fabric, which consists of two 4-input LUTs, two flip-flops (FFs) plus interconnecting circuitry. While this implementation fits comfortably into the Virtex-5 device, our aim is to produce a lower cost solution than a Virtex-5 implementation can provide.

Table-2: VIRTEX-5 FPGA Resource Usage Summary

Resource Type	Used	Available	Usage (%)
Slice registers	7381	28,800	25%
Total memory	234 KB	2160 KB	10%
BRAM	8	60	13%
DSP48	18	48	37%

5. Conclusion

The presented technique enables pitch scaling of speech by applying a simple linear translation of the CQT representation followed by a phase update stage to preserve phase coherence. It's computational less expensive, Informal listening test, however, suggest that these errors do not impair the perceived quality of the transposed output signal. That is, pitch transpositions in the CQT domain can be implemented very efficiently without performing frequency estimations.

6. Future Scopes

- To overcome the problems that arise for sharp transients in lower frequency areas several approaches have been outlined. The incorporation of one of the following techniques (or a combination of them) in the CQT phase vocoder implementation is an open issue:

i) Phase realignment for transients

ii) Transient separation: the input signal is split into a percussive and a harmonic part percussive events are then translated in time and added to the time-scaled harmonic part.

iii) Limiting the window lengths: To avoid very long windows in the first place, the window lengths could be limited towards low frequencies.

- Pitch transposition: In order to obtain a fully functional pitch-shifting algorithm based on the CQT, a formant preservation technique needs to be implemented. As for the CQT phase vocoder, phase coherence could be further improved by incorporating sinusoidal trajectory heuristics and shape invariant approaches.

- Note selective transpositions: The ease of transposing single notes in the CQT representation of polyphonic music signals has been demonstrated. However, the implementation of a piece of software that facilitates note selective transpositions in the CQT representation is a topic for future work.

References

[1] Christian Schörkhuber,AnssiKlapuri, AloisSontacchi "*Pitch Shifting Of Audio Signals Using The Constant-Q Transform*", Proc. of the 15th Digital Audio Effects (DAFx- 12), York, UK , September 17-21, 2012 Int. Conference.

[2] Shaikh Shafee,B.Anuradha "*Voice Conversion Using Different Pitch Shifting Approach over TD-PSOLA Algorithm"* International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue 12, December 2013

[3] Alexander G. Sklar, "*A Wavelet-based Pitch-shifting Method*", Filter Banks And Wavelets Een698.,2011.

[4] IlkerBayram, "*An Analytic Wavelet Traznsform wi th a Flexible Time- Frequency .Covering*", European Journal of Scientific Research, vol. 39, February 2011, pp. 309-315.

[5] Allam Mousa "*Voice Conversion Using Pitch Shifting Algorithm By Time Stretching with Psola And Re–Sampling* ''Journal of Electrical Engineering, VOL. 61, NO. 12010, 57– 61".

[6] Habib Estephan, Scott Sawyer, Daniel Wanninger, "*Real-Time Speech Pitch Shifting on an FPGA*", Department ofElectrical and Computer Engineering, Villanova University February 23, 2006.CA:University Science,1989.

[7] Magnus Erik Hvass Pedersen *"The Phase Vocoder and its Realization"* Daimi, University of Aarhus, May 2003.

[8] Christian Sch¨orkhuber, Anssi Klapuri1 "*Constant-Q Transform Toolbox For Music Processing"* 2011.

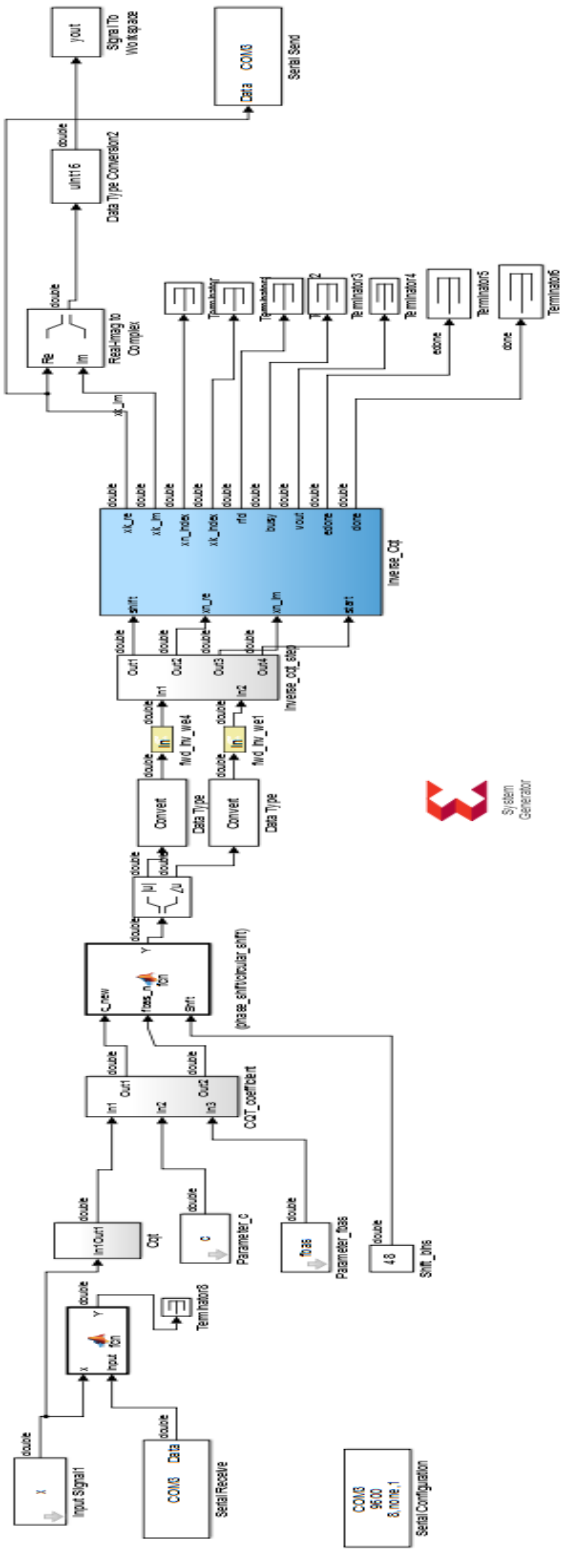
[9] www.mathworks.com

[10] www.xilinx.com – Datasheets Virtex-5 FPGA Configuration User Guide System Generator for DSP *Getting Started Guide ML505/506/507 Base System Builder Design Creation* AC '97 SoundMAX® Codec: AD1981B *EDK Concepts, Tools, and Techniques-A Hands-On Guide To Effective Embedded System Design.*

Author's Profile

Pranita R. Morbale has received her Master of Engineering (M.E.) degree with specialization of VLSI and Embedded Systems from Smt. Kashibai Navale College of Engineering, Pune in the year 2014. Her area of interest is audio processing VLSI and image processing.

Mahesh Navale has received his M.E in Electronics & Communication Engineering from Walchand College of engineering, Sangali. In the year 1999. At present he is working as an Assistant Professor at Sri Smt Kashibai Navale College of engineering, Pune. He is Research Research Scholar at Karpagam University, Coimbatore. His areas of interests are MATLAB, and Mobile networks.



Appendix I