ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol.2, Issue 12, December 2013

Voice Conversion Using Different Pitch Shifting Approach over TD-PSOLA Algorithm

Shaik Shafee[1], B. Anuradha[2]

Research Student, Department of ECE, SV University College of Engineering, Tirupati, India [1]

Professor, Department of ECE, SV University College of Engineering, Tirupati, India [2]

Abstract: Voice conversion and voice morphing has numerous applications in the commercial and industrial sectors. This paper emphasizes voice conversion using simple approach of pitch shifting method over TD-POSLA and re-sampling which depends on time stretching or time expanding based on the required target pitch. This study is performed to see the effect of voice conversion when some Telugu/Hindi (alphabets) speech signal is considered. Treatment of certain Telugu/Hindi voiced speech and vowels and the conversion between male, female and child speech has shown some expansion or compression in the resulting speech. The New approach is compared in terms of pitch shifting and computation/processing time with the conventional Time stretching and re-sampling algorithm is presented here. Analysis was performed for both male and female voice of Telugu/Hindi (alphabets) speech.

Keywords: vvoice conversion, voice morphing, PSOLA, re-sampling, pitch shifting, Telugu/Hindi alphabets, speech signals.

I. INTRODUCTION

Voice conversion is a process of modifying the source (speaker) voice to another speaker (target) voice. There are two major steps involving in conversion of voice. The first step is to identify the source voice parameters and the second is to apply these parameters to the required target voice. In this paper the major parameter Pitch is being taken for voice conversion. Speech is generally characterized by voiced which has high energy and high adjacent sample correlation (quasi-periodic wave) and unvoiced is characterized by a random aperiodic wave-form with low energy and low correlation.

There are number of time and frequency domain methods available for detecting the pitch of speech signal. This paper mainly focuses on changing the pitch of speech signal by different simple approach over conventional TD-PSOLA algorithm. Pitch modification or changing means transposing the pitch without changing the characteristics of the sound.

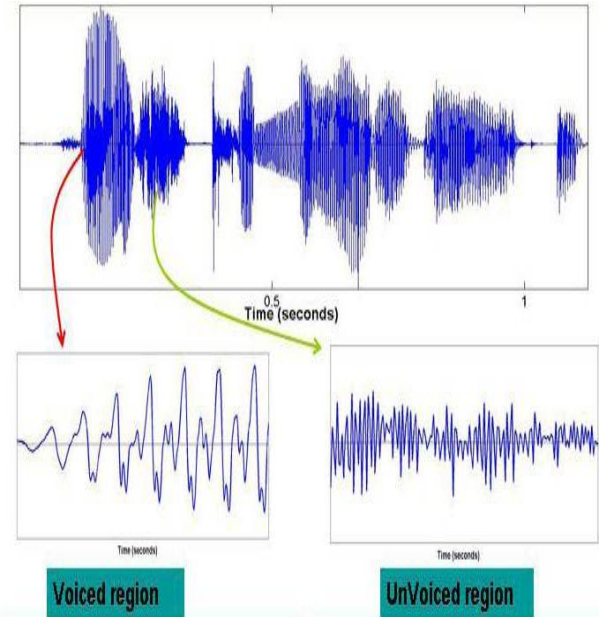


Fig.1: Waveform of a portion of: a) a voiced phone b) an unvoiced phone of speech signal.

II. PITCH PERIOD AND PITCH FREQUENCY

The voiced part of speech looks like a near periodic signal (quasi-periodic) in the time domain representation. In a short term, we may treat the voiced speech segments to be periodic for all practical analysis and processing. The periodicity associated with such segments is defined is 'pitch period ‘To' in the time domain and 'Pitch frequency or Fundamental Frequency fo' in the frequency domain. The term 'pitch' refers to the fundamental frequency ' fo'. Pitch is an important attribute of voiced speech. It contains speaker-specific information. It is also needed for speech coding task. The pitch can be estimated through different procedures like Autocorrelation method, Cepstrum Pitch determination, Pitch Estimate by SIFT method.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol.2, Issue 12, December 2013

Pitch shifting is a technique in which the original pitch of a sound is raised or lowered. The pitch period is responsible for making some sounds to be sharper than others. The aim of pitch shifting algorithms is to create a change in pitch without creating a change in the replay rate.

TABLE I

RANGE OF PITCH PERIOD

Speech by	Average (msec)	Maximum (msec)	Minimum (msec)
Male	8	12.5	5
Female	4.5	7	3
Child	3.5	5	2

TABLE II

RANGE OF PITCH FREQUENCY

Speech by	Average (Hz)	Maximum (Hz)	Minimum (Hz)
Male	125	80	200
Female	225	150	350
Child	300	200	500

Speech signals generally processed frame-by-frame. These short analysis segments, or "analysis frames" often overlap one another. Generally frame size will be taken as 10 ms to 15 ms for speech processing.

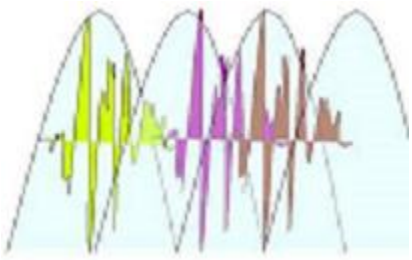


Fig 1: Windowing and framing of voiced speech

III. PSOLA (PITCH SYNCHRONOUS OVERLAP ADD)

It refers to a family of signal processing techniques that are used to perform time-scale and pitch-scale modification of speech .PSOLA is a method based on decomposition of a signal into a series of elementary waveforms in such a way that each waveform represents one of the successive pitch periods of the signal and the sum (overlap-add) of them reconstitutes the signal. PSOLA works directly on the signal waveform without any sort of model and therefore does not lose any detail of the signal. TD-PSOLA is the most commonly used due to its computational efficiency.

Pitch Shifting by Time Stretching and Re-sampling

In this algorithm the time duration of target speech will be maintained as same as source speech. This algorithm mainly consists of three following steps:

1. In analysis step, the original speech signal is divided into segments called as frames but often overlapping short-term analysis signals (ST). Short term signals Sn(k) are obtained from the discrete speech waveform s(n)by multiplying the signal by a sequence of the pitch-synchronous analysis window hn(k) .The window generally be of hamming type.

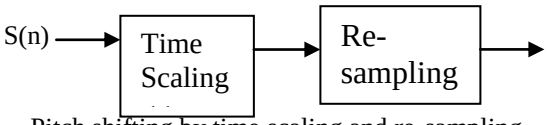
$$Sn(k) = hn(m - k)*S(n) \text{ -------(1)}$$

Where 'm' is an index for the short-time signal

2. The windows centred on the successive instants Tm, called pitch-marks. These marks are set at a pitch-synchronous rate on the voiced parts of the signal and at a constant rate on the unvoiced parts.

3. In the modification step, where each frame is modified according to the target. The synthesis steps are performed such that these segments are recombined by means of overlap adding.

The order of pitch shifting and times scaling can be changed as shown in below block diagram.



Pitch shifting by time scaling and re-sampling.

First, a time scaling algorithm expands the input signal from length N1 to length N2.Then a re-sampling operation with the inverse ratio N1/N2 performs pitch shifting and a reduction of length N2 back to length N1.

IV. DIFFERENT APPROACH IN MODIFICATION STEP

Rather than scaling and the re-sampling which involves bit complex with two steps, the following approach would modify the pitch (A) of source signal to another pitch (B) of target speech in single simple step. Each frame of frame size 'f' consists of k samples will be modified to same k samples with newer values based on pitch expansion or compression. If 'δ' is the ratio of target pitch (B) and source pitch (A) then 'δ' can be re written as:

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol.2, Issue 12, December 2013

δ = integer (i) +decimal (d) (2)
('d' is rounded to first order decimal.)

Then the new sample values of the frame of size 'f' can be calculated as :

$S_{new}(k) = S_{old}(\delta)$ (3)
(Where k=1, 2,3,…,frame size)

$S_{new}(k) = S_{old}(i+d)$ (4)

$S_{new}(k) \approx S_{old}(i) + d*[S_{old}(i+1) - S_{old}(i)]$ (5)
(for $\delta * k <= f$)

$S_{new}(k) \approx S_{old}(i-f) + d*[S_{old}(i+1-f) - S_{old}(i-f)]$ (6)
(for $\delta * k >= f$)

V. SIMULATION RESULTS

Both conventional i.e. Time stretching and re-sampling (PSOLA) method and new approach of pitch shifting techniques were analyzed on both male and female voice of Telugu/Hindi alphabets for different pitch ratios in MATLAB environment and respective processing times have been tabulated below for analysis.

TABLE III

PROCESSING TIME COMPARISON OF MALE VOICE

Alphabet (Male voice) (Telugu/Hindi) (44.1khz)	Pitchshift ratio (0.75) Process time(sec)		Pitchshift ratio (1.5) Process time(sec)		Pitchshift ratio (2.0) Process time(sec)	
	Method-1	Method-2	Method-1	Method-2	Method-1	Method-2
Am.wav అం/अम	0.094	0.024	0.098	0.026	0.101	0.027
Gha.wavఘ / घ	0.088	0.023	0.092	0.025	0.093	0.026
Ini.wav ఇని/ ईनी	0.092	0.032	0.098	0.023	0.098	0.021
Ra.wav ఠ / र	0.161	0.034	0.118	0.031	0.121	0.029

TABLE II

PROCESSING TIME COMPARISON OF MALE VOICE

Alphabet (Female voice (Telugu/Hindi) (44.1khz)	Pitchshift ratio (0.5) Process time(sec)		Pitchshift ratio (1.25) Process time(sec)		Pitchshift ratio (1.75) Process time(sec)	
	Method-1	Method-2	Method-1	Method-2	Method-1	Method-2
Am.wav అం/अम	0.086	0.021	0.089	0.018	0.091	0.016
Gha.wav ఘ/ घ	0.099	0.026	0.103	0.023	0.106	0.024
Ini.wav ఇని/ ईनी	0.088	0.029	0.091	0.019	0.096	0.017
Ra.wav ఠ / र	0.092	0.033	0.098	0.023	0.110	0.020

The processing time and pitch effect on changing the pitch of both male and female voices of Telugu/Hindi alphabets were analysed in MATLAB under the same configuration/environment. Method-1 assumed to be Conventional and Method-2 assumed to be New Approach in the the Tables.

The time and frequency response of Am.wav (male voice) and gha.wav (female voice) for different pitch shift ratio (δ) to achieve male to female/child and vice a versa presented here in in Fig-2 and Fig-3.

A. *Am.wav (Male voice)*

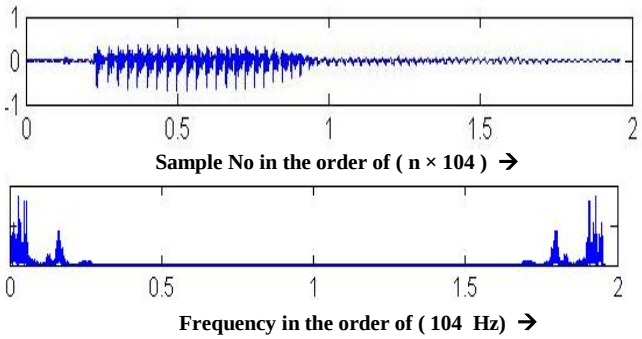
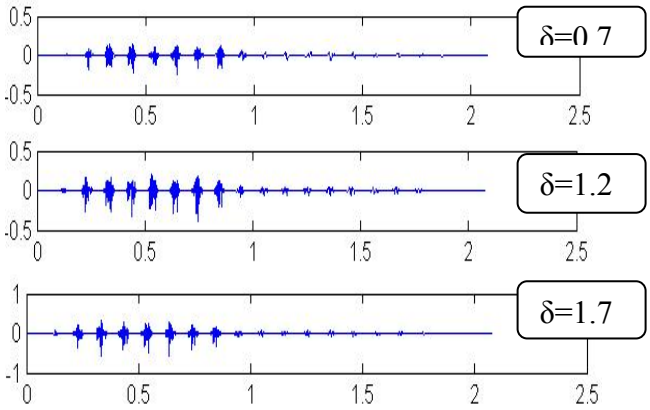


Fig 2(a): Time and frequncy response of Am.wav of 44.1khz speech sample

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol.2, Issue 12, December 2013

1) Conventional Method:



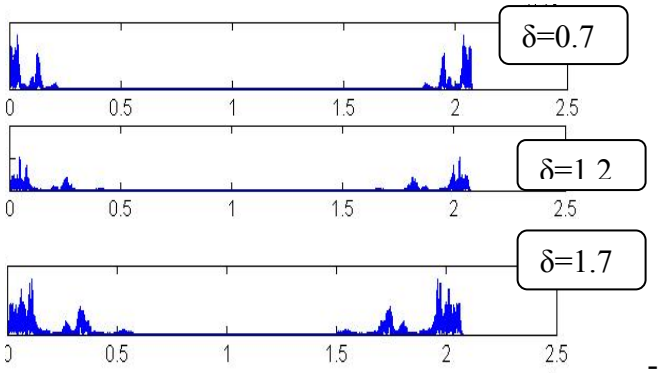
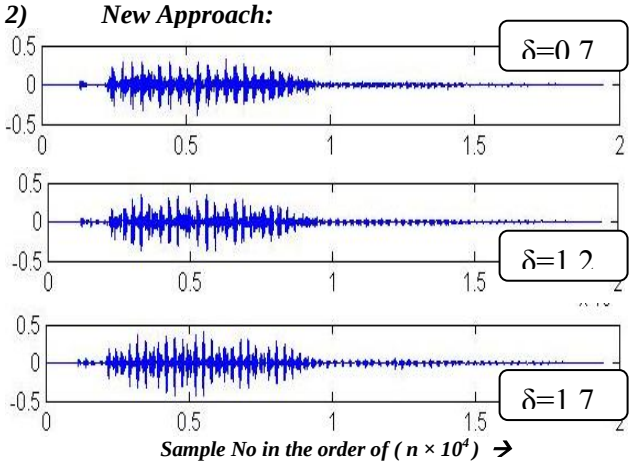


Fig 2(b): Time and frequncy response (FFT) of 'Am.wav' of 44.1 KHz speech sample in conventional method.

2) New Approach:



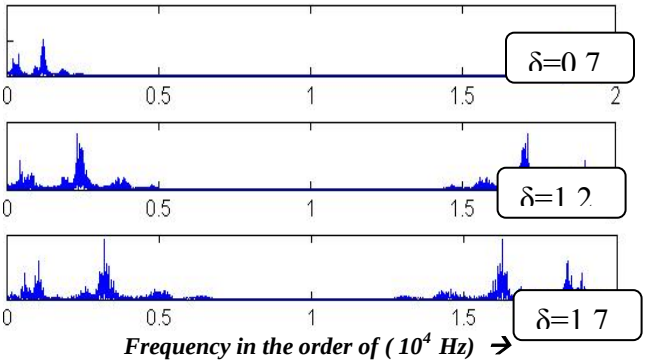


Fig 2(c): Time and frequncy response (FFT) of 'Am.wav' of 44.1 KHz speech sample in new approach method.

B. Gha.wav (Female voice)

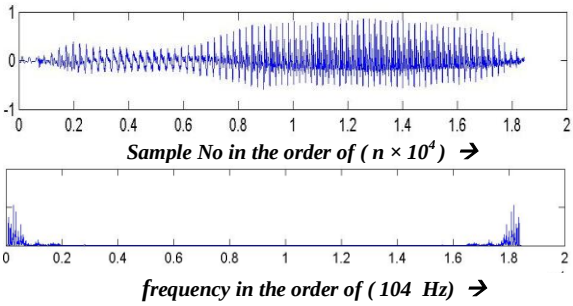
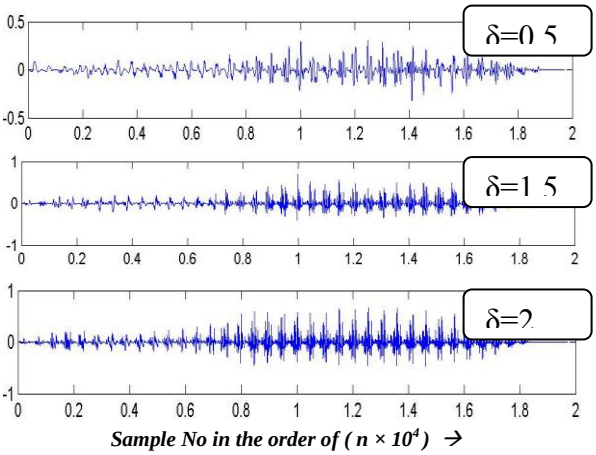


Fig 3(a): Time and frequncy response (FFT) of gha.wav of 44.1 Khz speech sample

1) Conventional Method:



ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol.2, Issue 12, December 2013

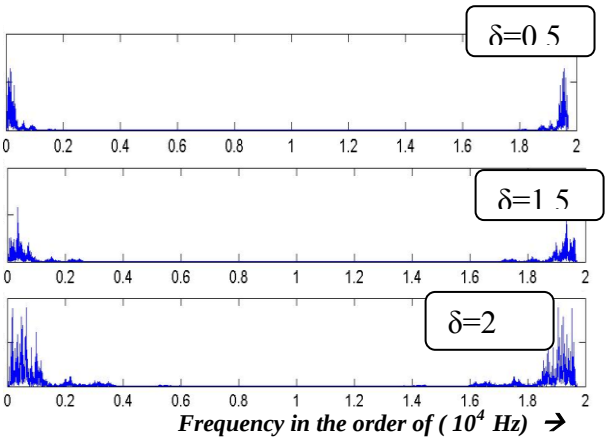


Fig 3(b): Time and frequncy response (FFT) of 'Gha.wav' of 44.1 KHz speech sample in conentional method.

2) New Approach:

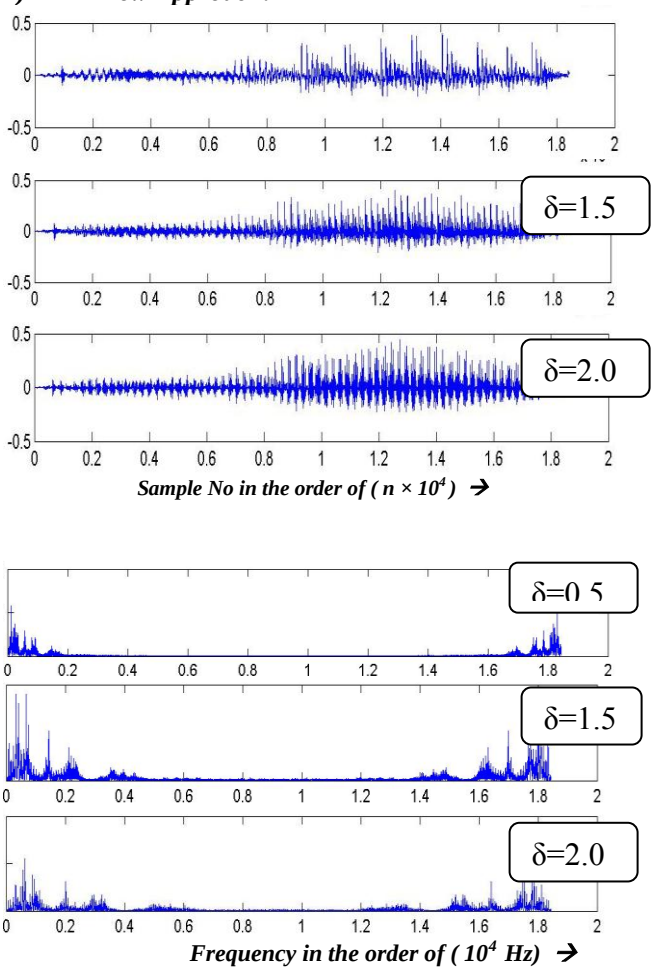


Fig 3(c): Time and frequncy response (FFT) of 'Gha.wav' of 44.1 KHz speech sample in new approach method

VI. DISCUSSIONS AND CONCLUSION

Different approach of Voice conversion depending on pitch-shifting over Time stretching using PSOLA and re-sampling was discussed. The processing time and pitch effect of Telugu/Hindi alphabets of recorded speech of both Male/Female voices are converted to female/Male/child and respective processing times were analyzed. The effect of pitch shift and processing time are much satisfactory in new approach. Further work can be extended on different approaches over LP-PSOLA and F-POSAL techniques.

REFERENCES

[1] A TD-PSOLA Based Method for Speech Synthesis and Compression, 978-1-4244-6363-3/10/$26.00c 2010 IEEE

[2] Allam Mousa: VOICE CONVERSION USING PITCH SHIFTING ALGORITHM BY TIME STRETCHINGWITH PSOLA AND RE–SAMPLING, Journal of ELECTRICAL ENGINEERING, VOL. 61, NO. 1, 2010, 57–61

[3] DAFX: Digital Audio Effects, Udo Zolzer, Copyright q 2002 John Wiley & Sons, Ltd.

[4] KONDOZ, A. M. : Digital Speech, Coding for Low Bit Rate Communication Systems, Wiley, 2004.