IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

# High Speed FFT Based Audio MORPHING Processor Using VHDL

**Palak Chawda [*1], Prof. Jaikaran Singh[2], Prof. Mukesh Tiwari [3]**

palak_shubhi7@yahoo.com

## Abstract

In this paper we form the voice morphing. We attempt to transform the spectral characteristics of a source speaker's speech signal so that the listener would believe that the speech was uttered by a target speaker. Voice morphing means the transition of one speech signal into another. Like image morphing, speech morphing aims to preserve the shared characteristics of the starting and final signals, while generating a smooth transition between them. For performing voice morphing we take a source voice and a targeted voice after applying FFT to extract the feature difference and store it in RAM then for morphing FFT source voice is applied to it and feature different is applied on it.

**Keywords**: Voice morphing, FFT

## Introduction

Voice morphing or voice conversion means the transition of one speech signal (Source) into another (target) while preserving the original meaning. Voice morphing technology has numerous applications such as text-to-speech adaptation, where the voice morphing system can be trained on relatively small amounts of data and allows new voices to be created at a much lower cost than the currently existing systems. The voice morphing system can also be used in the situation where the speaker was not available and previous recordings had to be used. Other applications include voice disguise as well as low bandwidth speech encoding, where speech may be transmitted without revealing the speaker's identity and then re-synthesized.

Like image morphing, speech morphing aims to preserve the shared characteristics of the starting and final signals, while generating a smooth transition between them. Speech morphing is analogous to image morphing. In image morphing the in-between images all show one face smoothly changing its shape and texture until it turns into the target face. It is this feature that a speech morph should possess. One speech signal should smoothly change into another, keeping the shared characteristics of the starting and ending signals but smoothly changing the other properties. The major properties of concern as far as a speech signal is concerned are its pitch and envelope information.

Since the 1990s, many techniques for voice conversion have been proposed [1-6].

The first approaches were based around linear predictive coding (LPC) [14].Where the residual error was measured and used to produce the excitation signal[3, 1, and 4].Most authors developed methods based on either the interpolation of speech parameters and modelling the speech signals using formant frequencies [2], Linear Prediction Coding (LPC) cepstrum coefficients [7],Line Spectral Frequencies (LSFs) [8], and harmonic-plus-noise model parameters[9] or based on mixed time- and frequency- domain methods to alter the pitch, duration, and spectral features. These methods are forms of single-scale morphing.

## Methodology

Speech morphing can be achieved by transforming the signal's representation from the acoustic waveform obtained by sampling of the analog signal, with which many people are familiar with, to another representation. To prepare the signal for the transformation, it is split into a number of 'frames' - sections of the waveform. The transformation is then applied to each frame of the signal. This provides another way of viewing the signal information. The new representation (said to be in the frequency domain) describes the average energy present at each frequency band.

Further analysis enables two pieces of information to be obtained: pitch information and the overall envelope of the sound. A key element in the morphing is the manipulation of the pitch information. If two signals with different pitches were simply cross-faded it is highly likely that two separate sounds will be heard. This occurs because

the signal will have two distinct pitches causing the auditory system to perceive two different objects.

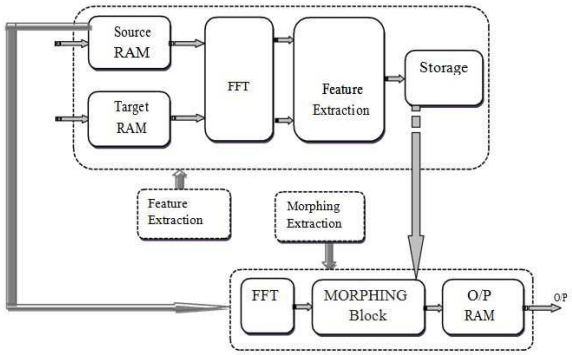



**Figure1: Block diagram of voice Morphing**

A successful morph must exhibit a smoothly changing pitch throughout. The pitch information of each sound is compared to provide the best match between the two signals' pitches. To do this match, the signals are stretched and compressed so that important sections of each signal match in time. The interpolation of the two sounds can then be performed which creates the intermediate sounds in the morph. The final stage is then to convert the frames back into a normal waveform.

So as per our work, we first take samples of source and destination voice and converted both to frequency domain and extracted features of both. Then difference between both is calculated, it gives info that how source is different from target. Then full source voice is varied by that difference thus we get morphed voice to target.

**Pitch detection**

Voiced speech signals can be considered as quasi-periodic. The basic period is called the pitch period. The average pitch frequency (in short, the pitch), time pattern, gain, and fluctuation change from one individual speaker to another. For speech signal analysis, and especially for synthesis, being able to identify the pitch is extremely important.

$$T_0 = \arg\max(\rho_z);\ \rho_z = \frac{< x, y >}{||x||.||y||};\ ||x|| = (< x, y >)^{1/2}$$

$$< x, y >= \int_{t_0}^{t_0+z} x(t)y(t)dt \ ;\ y(t) = x(t-\tau)$$

It is based on the fact that two consecutive pitch cycles have a high cross-correlation value, as opposed to two consecutive speech fractions of the same length but different from the pitch cycle time.

The pitch detector's algorithm can be given by equations

**Fourier Transform**

The Fourier transform (FT) of the function *f(x)* is the function F(ω), where:

$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx$$

$$f(x) = 1/2\pi \int_{-\infty}^{\infty} F(\omega)e^{i\omega x}d\omega$$

$$\text{Where } i = \sqrt{-1} \text{ and } e^{i\theta} = \cos\theta + i\sin\theta$$

Think of it as a transformation into a different set of basic functions. The Fourier transform uses complex exponentials (sinusoids) of various frequencies as its basis functions.

A Fourier transform pair is often written function $f(x) \leftrightarrow F(\omega)$, or $F(f(x)) = F(\omega)$ where F is the Fourier transform operator.

The Fast Fourier Transform (FFT) Algorithm.

There are many variants of the FFT algorithm.

We'll discuss one of them, the "decimation-in-time" FFT algorithm for sequences whose length is a power of two

Below is a diagram of an 8-point FFT, $W = W_8 = e^{i\pi/4} = (1-i)/\sqrt{2}$:

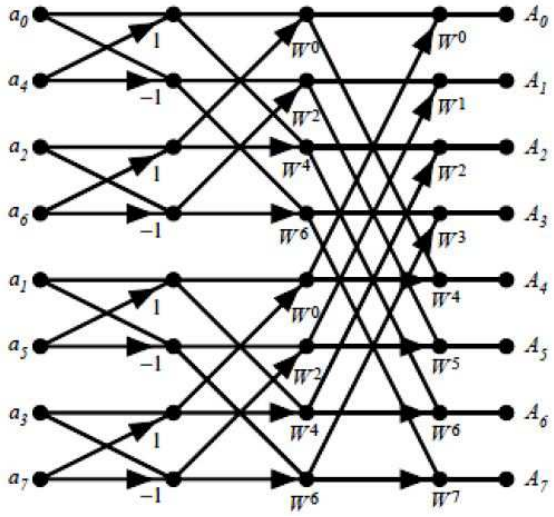
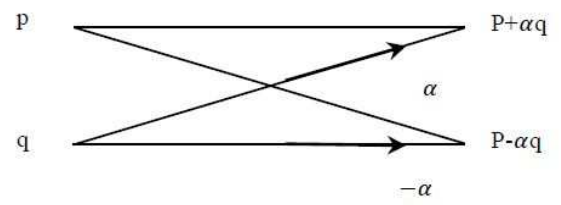



**Figure2: Butterflies diagram for 8-point FFT**

Each butterfly takes two complex numbers p and q and computes from them two other numbers, p + αq and p - αq, where is a complex number. Below is a diagram of a butterfly operation.




## Simulation And Results

Simulation is performed using Modelsim and verified by MATLAB. Figure below shows simulation waveform:

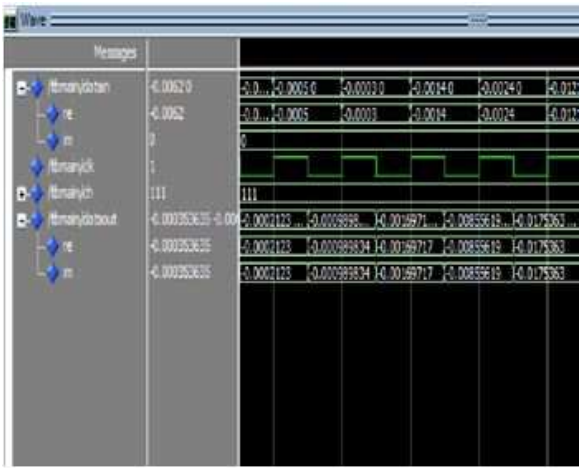


**Figure3: Simulation of voice morphing**

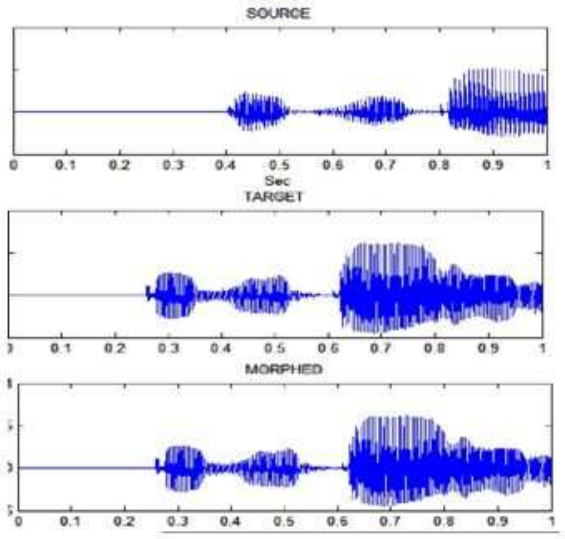



**Figure4: Simulation in MATLAB**

## Conclusion

There are basically three inter-dependent issues that must be solved before building a voice morphing system. Firstly, it is important to develop a mathematical model to represent the speech signal so that the synthetic speech can be regenerated and prosody can be manipulated without arti-facts. Secondly, the various acoustic cues which enable humans to identify speakers must be identified and extracted. Thirdly, the type of conversion function and the method of training and applying the conversion function must be decided. Here we presented a very simple scheme to produce voice morphing. Pitch of source is converted target using the information for feature difference. Simulation shows that it needed a bit more hardware. So other technique can be used to reduce hardware like phase vocoder.

## References

[1] L.M. Arslan, D.Talkin,"Voice conversion by codebook map-ping of line spectral frequencies and excitation spectrum," *Proc. Eurospeech*, pp.1347-1350, 1997. excitation spectrum," Proc. Eurospeech, pp.1347-1350, 1997.

[2] M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara: Voice conversion throughvector quantization. IEEE Proceedings of the IEEE ICASSP, 1998, 565–568.

[3] L. Arslan: Speaker transformation algorithm using segmental codebooks (stasc).Speech Communication 28, 1999, 211–226.

[4] Y. Stylianou, O. Cappe, and E. Moulines: Statistical methods for voice qualitytransformation. Proc. EUROSPEECH, 1995, 447–450.

[5] http://www.seminarpaper.com/2011/12/voice-morphing-full-report.html.

[6] Z. Shuang, F. Meng, Y. Qin, "Voice Conversion by Combining Frequency Warping with Unit Selection", in Proc.ICASSP, pp.4661-4664, 2008.

[7] S. Furui: Research on individuality features in speech waves and automaticspeaker recognition techniques. Speech Communication 5, 1986, 183–197.

[8] A. Kain and M.W.Macon: Spectral voice conversion for text-to-speech synthesis.Proc. ICASSP'98 1, 1998.

[9] H. Valbret, E. Moulines, and J.P. Tubach: Voice transformation using psolatechnique. Speech Communication 11, 1992, 175–187.