IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 1, Issue 2, September 2012)

A VHDL Implementation of Voice Morphing using FFT

Palak Chawda
palak_shubhi7@yahoo.com

Prof. Jaikaran Singh
jksingh81@gmail.com

Prof.Mukesh Tiwari
mukeshtiari_79@yahoo.co.in

***Abstract* – in this paper we form the voice morphing. 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**

I. Introduction

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. 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. These two reside in a convolved form in a speech signal. Hence some efficient method for extracting each of these is necessary. We have adopted an uncomplicated approach namely cepstral analysis to do the same. Pitch and formant information in each signal is extracted using the cepstral approach. Necessary processing to obtain the morphed speech signal include methods like Cross fading of envelope information, Dynamic Time Warping to match the major signal features (pitch) and Signal Re-estimation to convert the morphed speech signal back into the acoustic waveform. Voice morphing technology enables a user to transform one person speech pattern into another person pattern with distinct characteristics, giving it a new identity while preserving the original content. Many ongoing projects will benefit from the development of a successful voice morphing technology: text-to-speech (TTS) adaptation with new voices being created at a much lower cost than the currently existing systems; broadcasting applications with appropriate voices being reproduced without the original speaker being present; voice editing applications with undesirable utterances being replaced with the desired ones; internet voice applications with e-mail readers and screen readers for the blind as well as computer and video game applications with game heroes speaking with desired voices. A complete voice morphing system incorporates a voice conversion algorithm, the necessary tools for pre- and post-processing, as well as analysis and testing. The processing tools include waveform editing, duration scaling as well as other necessary enhancements so that the resulting speech is of the highest quality and is perceived as the target speaker. In the last few years many papers have addressed the issue of voice morphing using different signal processing techniques [1-6]. Most methods developed were single-scale methods based on the interpolation of speech parameters and modeling of the speech signals using formant frequencies [7], Linear Prediction Coding Cepstrum coefficients [9], Line Spectral Frequencies [8] and harmonic-plus-noise model parameters [2]. Other methods are based on mixed time- and frequency-domain methods to alter the pitch, duration and spectral features. The methods suffer from absence of detailed information during the extraction of formant coefficients and the excitation signal which results in the limitation on accurate estimation of parameters as well as distortion caused during synthesis of target speech.

IJDACR
ISSN: 2319-4863
IJDACR
International Journal Of Digital Application & Contemporary Research
International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 1, Issue 2, September 2012)

II. 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

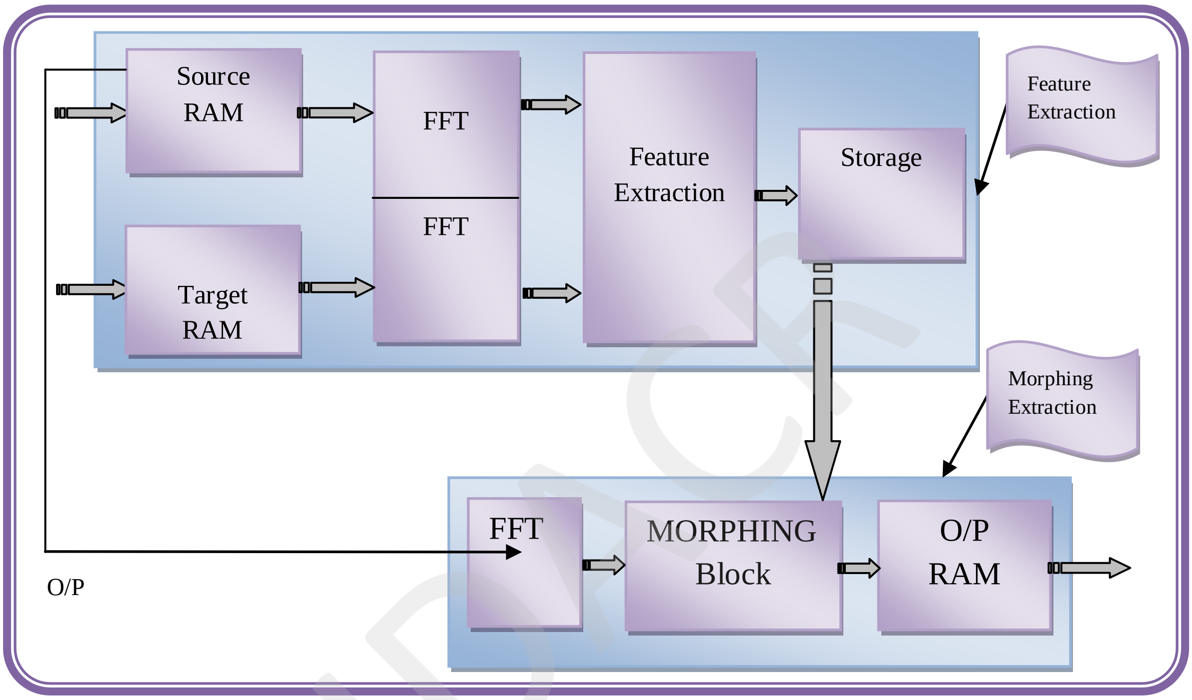

Figure.1 Block diagram of voice Morphing

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. 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.

However, after the morphing has been performed, the legacy of the earlier analysis becomes apparent. The conversion of the sound to a representation in which the pitch and spectral envelope can be separated loses some information. Therefore, this information has to be re-estimated for the morphed sound. This process obtains an acoustic waveform, which can then be stored or listened to.

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.

IJDACR
ISSN: 2319-4863

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 1, Issue 2, September 2012)

III. Simulation and Results

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

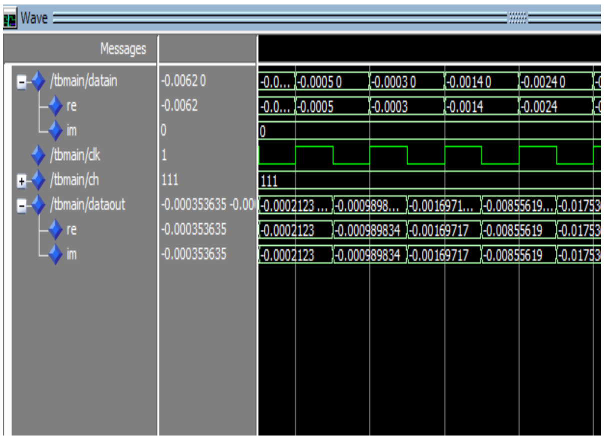

Figure.2. Simulation of voice morphing

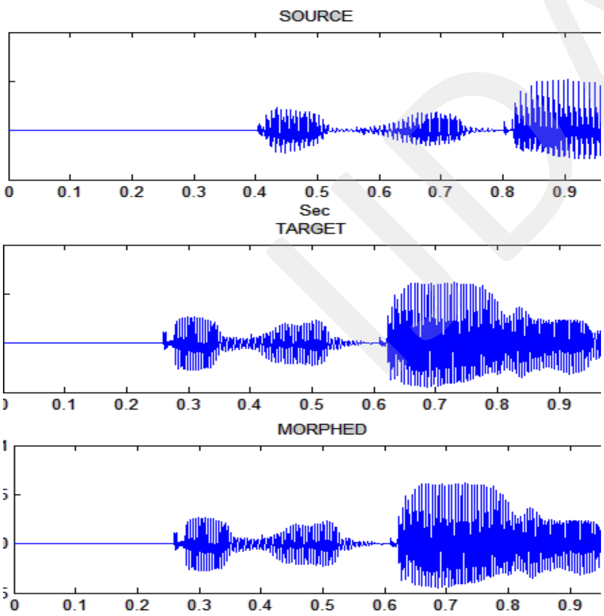

Figure.3. Simulation in MATLAB

IV. 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] D.H.Klatt "Review of Text-to-Speech Conversion for English" Journal of the Acoustical Society of America Vol.67.pp.971-995. 1980.

[2] D.H.Klatt "software for a cascade/parallel format synthesizer" Journal of the Acoustical Society of America **82**(3):737-93.

[3] T.Dutoit. "An introduction to Text-To-Speech in Android" kluwer.1997.

[4] R. Cole. Et al. Survey of the state of art in Human language technology. Chapter 5: Spoken Output technologies. Cambridge university press,1998.http://cslu.cse.ogi.edu/HLTsurvey/

[5] J. P. Olive, "the talking computer: text to speech synthesis, "in chapter 6 of HAL's Legacy D.G Stork. Mit press, 1996. http://mitpress.mit.edu/e-books/Hal/

[6] S. S. Agrawal. N. Pinto. R. Verma. A. S. Sarma and K.N. Stevens, "Synthesizing high Quality Hindi Speech Using KLSY88."Journal of Acoustic Society of India. Vol.20 No. 1-4, 1992.

[7] R. Verma, A. S. Sarma , K.N. Stevens, S. S. Agrawal, N. Shrotiya, "On the Development of Text-to speech System for Hindi," proceedings of ICPhS 95.Stockholm.1995.