IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 2, Issue 8, March 2014)

High Speed FFT based Audio MORPHING Processor using VHDL

Navneet Kour Gandhi
gandhinavneet.27@gmail.com
Prof. Preet Jain
preetjain@gmail.com

Abstract— **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, RAM.**

I. INTRODUCTION

Voice morphing is an area of speech processing that deals with the conversion of the perceived speaker identity. In other words, the speech signal uttered by a first speaker, the source speaker, is modified to sound as if it was spoken by a second speaker, referred to as the target speaker. 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; 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. Yet other possible applications could be speech-to-speech translation and dubbing of television programs.

Despite the increased research attention that the topic has attracted, voice morphing has remained a challenging area. One of the challenges is that the perception of the quality and the successfulness of the identity conversion are largely subjective. Furthermore, there is no unique correct conversion result.

One speech signal can be easily altered into another, preserving the shared features of the starting and ending signals but effortlessly changing the other properties. The key features of a speech signal are its pitch and envelope information. These two reside in a convolved form in a speech signal. Therefore some effective technique for mining each of these is essential.

Since the 1990s, many techniques for voice conversion have been proposed [5-11].

The first approaches were based around linear predictive coding (LPC).Where the residual error was measured and used to produce the excitation signal [7, 5, and 8]. Most authors developed methods based on either the interpolation of speech parameters and modelling the speech signals using formant frequencies [6], Linear Prediction Coding (LPC) cepstrum coefficients [11],Line Spectral Frequencies (LSFs) [12], and harmonic-plus-noise model parameters[13] 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.

Speech processing

Speech processing is the study of the speech signals and hence the processing methods of these signals. The signals may be in analogue or digital format but usually it is processed in a digital representation. Speech processing can be divided into several categories like [14, 15].

- Speech recognition, dealing with analysis of the linguistic content of a speech signal.
- Speaker recognition, aiming to recognize the identity of the speaker.
- Enhancement of speech signals and noise reduction.
- Speech coding, which is mainly used in data compression and telecommunications.
- Voice analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords.
- Speech synthesis to produce an artificial synthesis of speech like computer generated speech.

It is needed to analyze the speech signal into its components which are the excitation signal and the linear Prediction filter but before analysis the signal must be pre-emphasis using the pre-emphasis filter. Moreover, pitch determination usually has an important role in speech processing

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 2, Issue 8, March 2014)

Speech Morphing

The concept of morphing relies heavily upon the fact that specific algorithms can synthesize the various characteristics of voice. If one had two speakers "A" and "B", and we wanted to take what "A" said, but make it come out in "B's". In theory, one should take the excitation function of "A", map the pitch from the excitation function of "B" on to it using certain algorithms like the FFT algorithm, and then pass it through the filter created by the cavities and articulators of "B". This should synthesize "A" words using "B" pitch and formants [16].

The most important part of voice morphing is speech synthesis, since the quality of the synthesized speech is the ultimate aim of voice conversion. Speech signals will be synthesized by means of the same parametric representation that was used in the analysis. It can be synthesized from the linear predictive analysis parameters. A simplified flowchart showing the procedure used here to achieve speech morphing is illustrated in Figure 1.

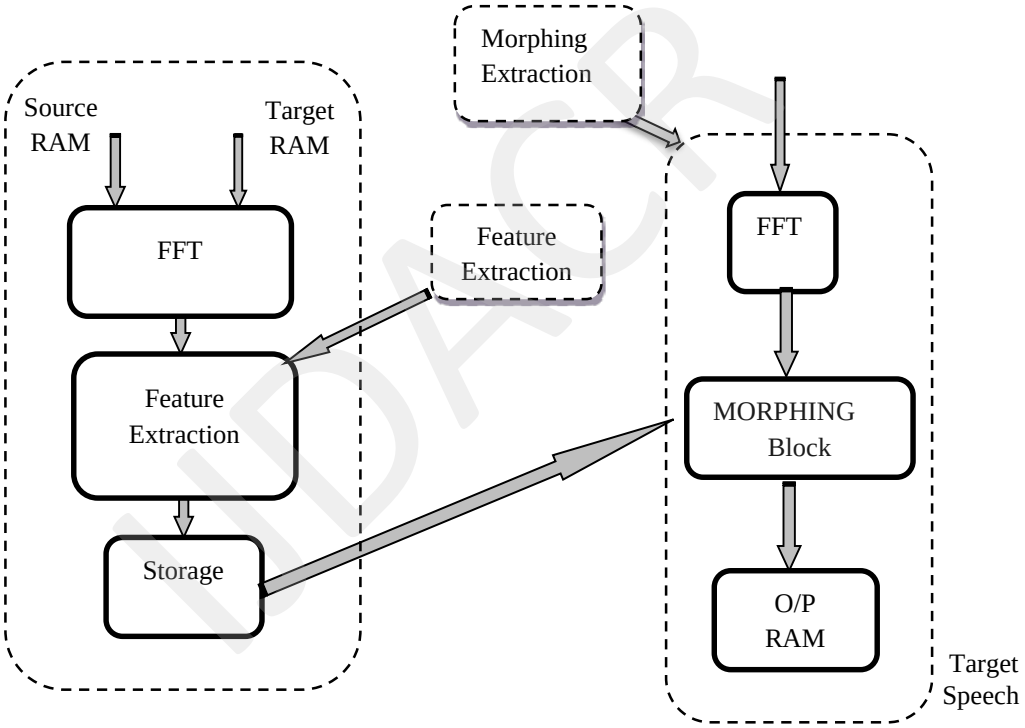

Figure 1. Flowchart of the speech morphing algorithm

II. METHODOLOGY

Voice morphing can be carried out by altering the signal's representation from the acoustic waveform received by sampling of the analog signal, with which many people are familiar with, to another representation. To organize the signal for the alteration, it is divided into a number of small 'frames'. Then each frame is treated with this alteration. The novel demonstration describes the average energy present at each frequency band.

This analysis contain two pieces of information to be obtained: the overall envelope of the sound and the pitch information. A key factor in the morphing approach is the manipulation of the pitch information.

The pitch data of each sound is associated to provide the best match between the two signals' pitches. To do this match, the signals are stretched and compressed so that the vital parts of each signal match in time. The interpolation of the two sounds can then be done which creates the intermediate sounds in the morph.

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 2, Issue 8, March 2014)

So as per our work, we first take samples of source and destination voice and converted both to frequency domain using FFT 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.

Fourier Transform

$$W_N^{k+n/2} = e^{-j\left(\frac{2\pi k}{N}+\pi\right)}$$

$$= e^{-j2\pi k/N}$$

$$= -W_N^k$$

Therefore,

$$X^{(N)}(k) = X_0^{\left(\frac{N}{2}\right)}(k) \mathrel{+}= W_N^k X_1(k)$$

$$for\ k = 0, \ldots, \frac{N}{2} - 1$$

$$X^{(N)}\left(k + \frac{N}{2}\right) = X_0^{\left(\frac{N}{2}\right)}(k) - W_N^k X_1(k)$$

$$for\ k = 0, \ldots, \frac{N}{2} - 1$$

$$X(0) = \sum_{n=0}^{0} x(n) e^{-j\left(\frac{2\pi.0}{1}\right)n} = x(0)$$

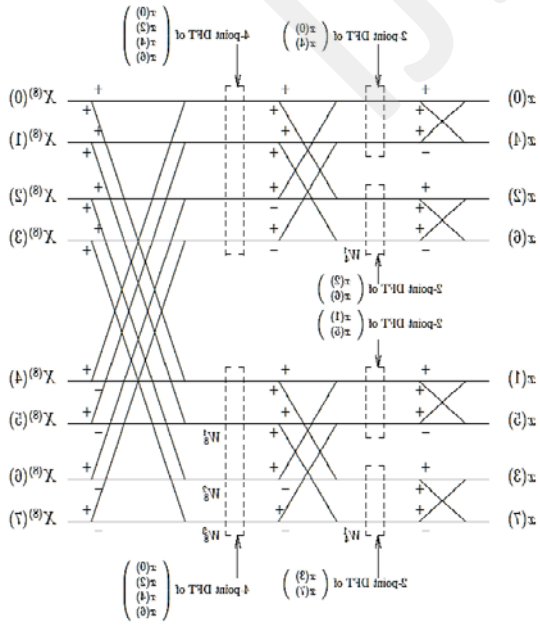

Figure 2.The 8-point FFT.

III. SIMULATION ANDRESULTS

* Advanced HDL Synthesis *

Loading device for application Rf_Device from file '3s500e.nph' in environment C:\Xilinx92i.

Advanced HDL Synthesis Report

Macro		Statistics
# Multipliers	:	240
32x32-bit multiplier	:	36
32x5-bit multiplier	:	48
33x33-bit multiplier	:	48
5x5-bit multiplier	:	108
# Adders/Subtractors	:	342
32-bit adder	:	24
32-bit adder carry out	:	84
32-bit subtractor	:	12
5-bit adder	:	86
5-bit subtractor	:	112
6-bit adder	:	24

Device utilization summary:

Selected Device: 3s500efg320-4
Number of Slices: 916 out of 4656 19%
Number of 4 input LUTs: 1517 out of 9312 16%
Number of IOs: 160
Number of bonded IOBs: 160 out of 232 68%
Number of MULT18X18SIOs: 7 out of 20 35%
Maximum combinational path delay: 60.176ns

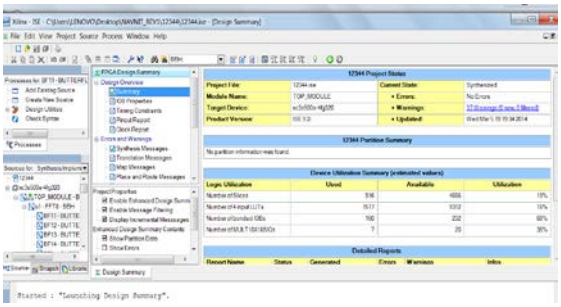

Figure 3. Device summary

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 2, Issue 8, March 2014)

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 artifacts. 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.

V. REFERENCES

1. Cano, P. "Fundamental Frequency Estimation in the SMS Analysis". Proceedings of the Digital Audio Effects Workshop, 1998.
2. Childers, D.G.. "Measuring and Modeling Vocal Source-Tract Interaction". IEEE Transactions on Biomedical Engineering 1994.
3. Cano, P., A. Loscos. Singing Voice Morphing System based on SMS. UPC, 1999.
4. 4. Cano, P., A. Loscos, J. Bonada, M. de Boer, X. Serra. 2000. "Singing Voice Impersonator Application for PC". Proceedings of the ICMC 2000.
5. L.M. Arslan, D.Talkin,"Voice conversion by codebook map-ping of line spectral frequencies and excitation spectrum," Proc. Eurospeech, pp.1347-1350, 1997.
6. M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara: Voice conversion through vector quantization. IEEE Proceedings of the IEEE ICASSP, 1998, 565–568.
7. L. Arslan: Speaker transformation algorithm using segmental codebooks (stasc).Speech Communication 28, 1999, 211–226.
8. Y. Stylianou, O. Cappe, and E. Moulines: Statistical methods for voice quality transformation. Proc. EUROSPEECH, 1995, 447–450.
9. http://www.seminarpaper.com/2011/12/voice-morphing-full-report.html.
10. Z. Shuang, F. Meng, Y. Qin, "Voice Conversion by Combining Frequency Warping with Unit Selection", in Proc.ICASSP, pp.4661-4664, 2008.
11. S. Furui: Research on individuality features in speech waves and automatic speaker recognition techniques. Speech Communication 5, 1986, 183–197.
12. A. Kain and M.W.Macon: Spectral voice conversion for text-to-speech synthesis. Proc. ICASSP'98 1, 1998.
13. H. Valbret, E. Moulines, and J.P. Tubach: Voice transformation using psola technique. Speech Communication 11, 1992, 175–187.
14. Kondoz M., Digital Speech, Coding for Low Bit Rate Communication Systems, Press Wiley, 2004.
15. Rabiner R. and Juang B., Fundamentals of Speech Recognition, Prentice Hall, 1993.
16. Pfitzinger R., "Unsupervised Speech Morphing between Utterances of any Speakers," in Proceedings of the 10th Australian International Conference on Speech Science and Technology, Sydney, pp. 8-10, 2004.