Paper from Proceeding of the National Conference on Innovative Developments in Science, Technology & Management (NCIDSTM-2015) Organized by Ganga Technical Campus, Soldha, Bahadurgarh, Haryana (India) March 1st 2015 Published by International Journal of Engineering Sciences Paradigms and Researches (IJESPR) with ISSN (Online): 2319-6564, Impact Factor: 2.20 and Website: www.ijesonline.com

VHDL Implementation of Phase Vocoder for Voice Morphing

Hitanshu Saluja[1], Ritu Juneja[2] and Kavinder Singh[3]
[1]Electronics & Communication Department, Ganga Technical Campus
Bahadurgarh, Haryana, India
hitanshuu@gmail.com

[2]Electronics & Communication Department, Ganga Technical Campus
Bahadurgarh, Haryana, India
rtjuneja@gmail.com

[3]CNS Department, Airport Authority of India
kavirao38@gmail.com

Abstract

Speech is usually characterized as voiced, unvoiced or transient forms. Voiced speech is produced by an air flow of pulses caused by the vibration of the vocal cords. The resulting signal could be described as quasi-periodic waveform with high energy and high adjacent sample correlation. The two broad categories of pitch-estimation algorithms are time-domain algorithms and frequency-domain algorithms.

Voice morphing is the process of producing intermediate or hybrid voices between the utterances of two speakers. It can also be defined as the process of gradually transforming the voice of one speaker to that of 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. This paper describes the Phase vocoder method of voice morphing. We extract the feature difference of source and targeted voice apply this phase different to the source voice.

Keywords: Phase Vocoder, Voice Morphing, Speakers, Vocal Cords.

I. INTRODUCTION

Voice morphing is the modification of a speaker voice called source speaker— in order to make it being perceived as if another speaker —target speaker— had uttered it. Given thus two speakers, the aim of a voice morphing system is to determine a transformation function that converts the speech of the source speaker (from which usually a complete database is available) into the speech of the target speaker (from which normally few data are available), replacing the physical characteristics of the voice without altering the message contained in the speech

Human voice imitation can be found in three main aspects of daily human communication: language acquisition, impersonation for entertainment, and voice disguise for concealing a personal identity. Nevertheless, human imitation is not the only way to imitate others' voices: automatic voice morphing is the modification of a source speaker in order to create the perception that it was uttered by target speaker. Given thus two speakers, the aim of a voice morphing system is to determine a transformation function (TF) that converts the speech of the source speaker into the speech of the target speaker, replacing the speaker characteristics of the voice without altering the message contained in the speech.

The development of speech technologies has led to a wide variety of research areas related to different tasks involved in making computers interact orally with humans: modelling of speech production and perception, prosody analysis and generation, speech and audio processing, enhancement, coding and transmission, speech synthesis, analysis and synthesis of emotions in expressive speech, speech and speaker recognition, speech understanding, accent and language identification, cross- and multi-ingual processing, multimodal signal processing, dialogue systems, information retrieval, translation, applications for handicapped persons, etc.

Voice morphing systems have to be capable of accomplishing two main tasks:

1) Given a certain amount of training data recorded from specific source and target speakers, the system has to determine the optimal transformation for converting one voice into the other one [16].

2) The system has to apply this optimal transformation to convert new input utterances of the source speaker [17].

II. MOTIVATION

Despite the increased research attention that the topic has attracted, voice conversion 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.

There are many applications of voice morphing including customizing voices for text to speech (TTS) systems, transforming voice-overs in adverts and films to sound like that of a well-known celebrity, and enhancing the speech of impaired speakers such as laryngectomees. Two key requirements of many of these applications are that firstly they should not rely on large amounts of parallel training data where both speakers recite identical texts, and secondly, the high audio quality of the source should be preserved in the transformed speech. The core process in a voice morphing system is the transformation of the spectral envelope of the source speaker to match that of the target speaker and various approaches have been proposed for doing this. Yet other possible applications could be speech-to-speech translation and dubbing of television programs.

III. OBJECTIVE

There are three interdependent issues that must be decided before building a voice morphing system. First, a mathematical model must be chosen which allows the speech signal to be manipulated and regenerated with minimum distortion. Previous research suggests that the sinusoidal model is a good candidate since, in principle at least, this model can support modifications to both the prosody and the spectral characteristics of the source signal without inducing significant artifacts However, in practice, conversion quality is always compromised by phase incoherency in the regenerated signal, and to minimize this problem, we use phase vocoder approach.

Second, the acoustic features which enable humans to identify speakers must be extracted and coded. These features should be independent of the message and the environment so that whatever and wherever the source speaker speaks, his/her voice characteristics can be successfully transformed to sound like the target speaker.

Third, the type of conversion function and the method of training and applying the conversion function must be decided. The aim of this work is to give a system overview of the voice morphing. This approach addresses theoretical and practical aspects, so MathWorks MATLAB® was chosen as an implementation and evaluation framework for the realization of an exemplary voice morphing system

IV. 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.

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.

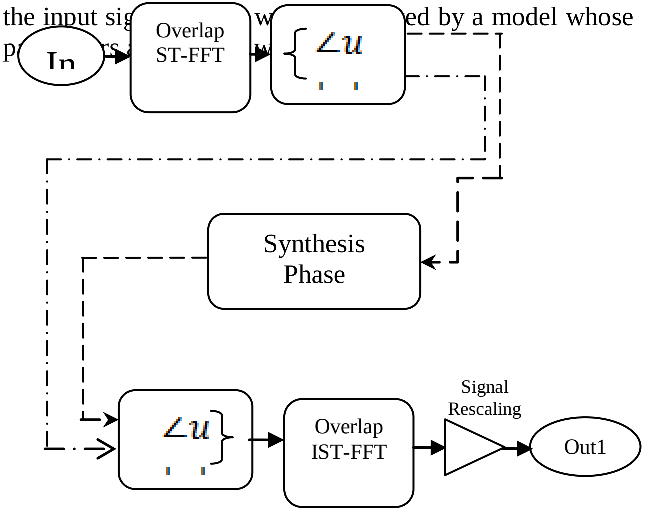
However, after the morphing has been performed, the legacy of the earlier analysis becomes apparent. The conversion of the sound to a representation in which the pitch and spectral envelope can be separated loses some information. Therefore, this information has to be re-estimated for the morphed sound.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

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.

The aim of voice morphing is to produce natural sounding hybrid voices between two speakers, uttering the same content. The algorithm is based on the concept phase vocoding, where the pitch of source voice is shifted by the phase difference between source and target Voice.

For composers interested in the modification of natural sounds, the phase vocoder is a digital signal processing technique of potentially great significance. By itself, the phase vocoder can perform very high fidelity time-scale modification or pitch transposition of a wide range of sounds. In conjunction with a standard software synthesis program, the phase vocoder can provide the composer with arbitrary control of individual harmonics. But use of the phase vocoder to date has been limited primarily to experts in digital signal processing.

Historically, the phase vocoder comes from a long line of voice coding techniques which were developed primarily for the electronic processing of speech. Indeed, the word ``vocoder'' is simply a contraction of the term ``voice coder.'' There are many different types of vocoders. The phase vocoder was first described in 1966 in an article by Flanagan and Golden. However, it is only in the past ten years that this technique has really become popular and well understood. The phase vocoder is one of a number of digital signal processing algorithms which can be categorized as analysis-synthesis techniques. Mathematically, these techniques are sophisticated algorithms which take an input signal and produce an output signal which is either identical to the input signal or a modified version of it. The underlying assumption is that the input sig w ed by a model whose p rs w



V. RESULTS

The program for FPGA is developed using the concept of phase vocoder in VHDL. The simulation is done on modelsim6.3f.

Figure 5.1: Simulation Waveform

Figure below showing the RTL of Phase Vocoder:

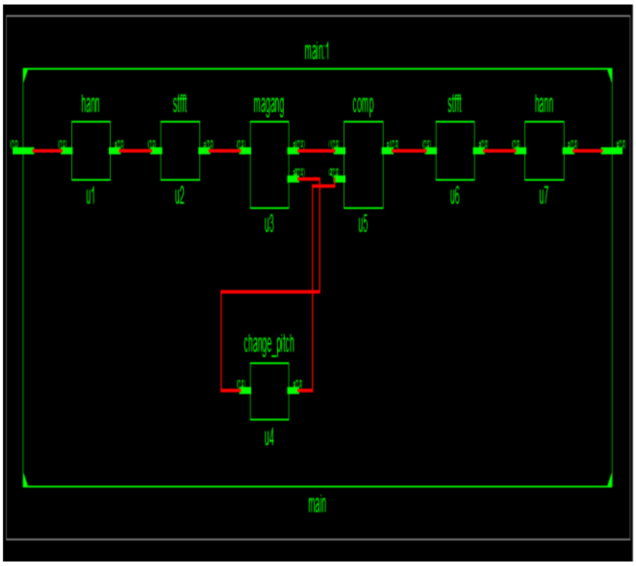


Figure 5.2: RTL of Phase Vocoder

References

[1]. "Analysis of a Modern Voice Morphing Approach using Gaussian Mixture Models for Laryngectomees", Aman Chadha, Bharatraaj Savardekar, Jay Padhya, International Journal of Computer Applications (0975 – 8887) Volume 49– No.21, July 2012.

[2]. Voice conversion: A critical survey, Anderson F. Machado, Marcelo Queiroz, 2010.

[3]. A shape-invariant phase vocoder for speech transformation, A. Robel, September 6-10, 2010.

[4]. Towards Morphological sound description using segmental models, Julien Bloit, Nicolas Rasamimanana, September 1-4, 2009.

[5]. A precise estimation of vocal tract parameters for high quality voice morphing, Ning Xu , Zhen Yang , 26-29 Oct. 2008.

[6]. Voice Conversion Based on Maximum Likelihood Estimation of Spectral Parameter Trajectory Tomoki Toda, Member, IEEE, Alan W Black, Member, IEEE, Keiichi Tokuda, Member, IEEE, August 1, 2007.

[7]. Ye H. and Young S., "High quality voice morphing", International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2004, Montreal, Vol. 1, 9-12.

[8]. C. Orphanidou, I. M. Moroz, and S. J. Roberts, "Wavelet based voice morphing," in WSEAS Transactions on Systems, December 2004, vol. 3, pp. 3297–3302.

[9]. Voice Morphing System for Impersonating in Karaoke Applications, Pedro Cano, Alex Loscos, Jordi Bonada, Maarten de Boer and Xavier Serra, 2000.

[10]. Speaker Transformation Algorithm using Segmental Codebooks (STASC) , Levent M. Arslan, 8 February 1999.

[11]. Automatic audio morphing, Malcolm Slaney, Michele Covell and Bud Lassiter, Presented at the 1996 International Conference on Acoustics,Speech, and Signal Processing, Atlanta.

[12]. Phase Vocoder, J.L.FLANAGAN and R.M.GOLDEN, July 18, 1966.

[13]. Voice transformation using PSOLA technique, H. Valbret, E. Moulines, J.P. Tubach, and June 1992.

[14]. M. Dolson, "The phase vocoder: A tutorial," Computer Music Journal, vol. 10, no. 4, pp. 14–27, 1986.

[15]. Erro, D., Moreno, A.: Sistema de s´ıntesis arm´onico/estoc´astico en modo pitchas ´ıncrono aplicado a conversi´on de voz. In: Proceedings of the IV Jornadas en Tecnolog´ıa de Habla. Zaragoza (2006)

[16]. Speaker Recognition Robustness to Voice Conversion, Mireia Farrus, Daniel Erro, and Javier Hernando.

[17]. Automatic speaker recognition as a measurement of voice imitation and conversion Mireia Farrús, Michael Wagner, Daniel Erro and Javier Hernando.

[18]. L. R. Rabiner , R. W. Schafer "Digital Processing of Speech Signals".

[19]. Yoav Meden, Eyal Yair and Dan Chazan, "Super Resolution Pitch Determination of Speech Signals", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 39, NO 1, JANUARY 1991.

[20]. McAulay, R. J., Quatieri, T. F., 1986. Speech analysis/synthesis based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech, and Signal Processing 34 (4), 744–754.

[21]. Stylianou. Modeling speech based on harmonic plus noise models. In Nonlinear Speech Modeling and Apps. Springer, 2005.

[22]. The Vocoder—Electrical Re-Creation of Speech, Homer Dudley, JSMPE; March 1940; 34 :(3) 272-278.

[23]. Speech Coding: A Thtorial Review, ANDREAS S. SPANIAS, MEMBER, IEEE.

[24]. Theory and Application of Digital Speech Processing, L. R. Rabiner and R. W. Schafer, June 3, 2009.

[25]. Phase Vocoder, J.I.FLANAGAN and R.M.GOLDEN, July 18, 1966.

[26]. An introduction to phase vocoder, John Gordon and John Strawn, Feb. 1987.

[27]. Harmonic Sinusoid Modeling of Tonal Music Events, Wen, Xue, A thesis submitted for the degree of Doctor of Philosophy of the University of London October, 2007.

[28]. Multiband excitation vocoder, D.W. Griffin, Aug. 1988.