IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 3, Issue 7, February 2015)

A PSOLA based Approach for Voice Morphing

Anant Bhatt
M. E. Scholar,
Poornima Group of Colleges, Rajasthan (India)
anantbhatt@poornima.org

Abstract – **Voice morphing is a name given to procedures which take speech as input from one speaker and attempt to generate speech that sounds like it came from another speaker. One compelling argument for good voice morphing is that it lessens the trouble in creating additional synthetic voices with new characters and styles once an existing voice has been created based on a full-sized corpus. There are further voice transformation applications for security, privacy, and assistive technologies. Although current voice transformation techniques perform well in the sense that humans typically judge transformed speech to sound more like the target speaker than the source speaker, there is still room for improvement. In this paper, we applied PSOLA for voice morphing as PSOLA works by dividing the speech waveform in small overlapping segments. To change the pitch of the signal, the segments are moved further apart (To decrease the pitch) or closer together (to increase the pitch). To change the duration of the signal, the segments are then repeated multiple times (To increase the duration) or some are eliminated (to decrease the duration). The segments are then combined using the overlap add technique. Results show that this work efficiently changes the pitch property of source to target.**

Keywords – **PSOLA, Voice morphing.**

I. INTRODUCTION

Among all the mechanisms that allow humans communicating and interacting with each other, speech is the most natural and precise one. Nowadays, the scientific community tries to face the challenge of designing speech-based human-computer interfaces, extending the role of speech to certain real-life situations in which more primitive ways of interaction (keyboard, mouse, joystick, graphic user interfaces, commands, buttons, etc.) are used to present. In other words, it is intended to make machines recognize well what human speakers say, and answer by generating output utterances that the listeners are capable of understanding, trying to imitate the human way of communicating with similar naturalness and precision. 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-lingual processing, multimodal signal processing, dialogue systems, information retrieval, translation, applications for handicapped persons, etc.

The voice morphing model should be capable of performing two main tasks:

- By using a certain amount of training information logged from a particular source and target speakers, the model has to conclude the optimal morphing information for converting one voice into the other one [1]
- The model has to apply this optimal morphing information to convert new input utterances of the source speaker [2].

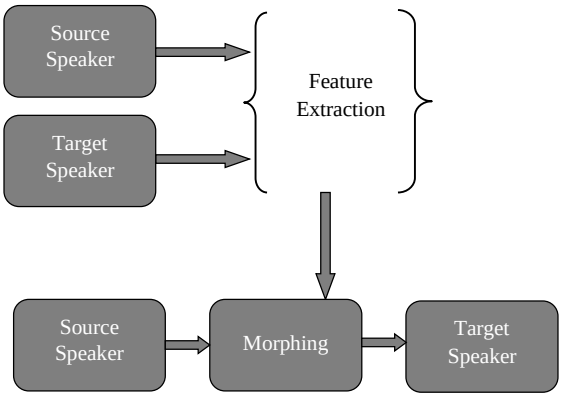

Figure 1: Basic flow of voice morphing

In this form of framework, speech synthesis can be considered as the artificial construction of human speech. The central topic of this thesis, voice morphing, can be considered a part of the speech synthesis area. The objective of voice conversion models is to transform the voice created by a specific

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 3, Issue 7, February 2015)

speaker, referred to as the source speaker, into a different particular speaker, named as target speaker [3]. Therefore, the features of the source speaker have to be recognized by the model and swapped with those of the target speaker, without dropping any information or enhancing the message that is being transmitted.

There are some issues that must be discussed before constructing a voice morphing model. Initially, an accurate model must be selected which permits the speech signal to be regenerated and manipulated with least alteration. The earlier study recommends that the sinusoidal system is a good candidate, in principle at least, this kind of system can support changes to both the prosody and the spectral features of the source signal without the use of significant artifacts. But, in run time, the conversion excellence is always negotiated by phase incoherency in the redeveloped signal, in order to diminish this problem, we use the phase vocoder approach.

Second, the acoustic information which allows humans to recognize speakers must be mined and coded. This information should not depend on the message and the environment so that wherever the source speaker speaks, his/her voice features can be positively transformed to sound like the target speaker.

II. PROPOSED METHOD

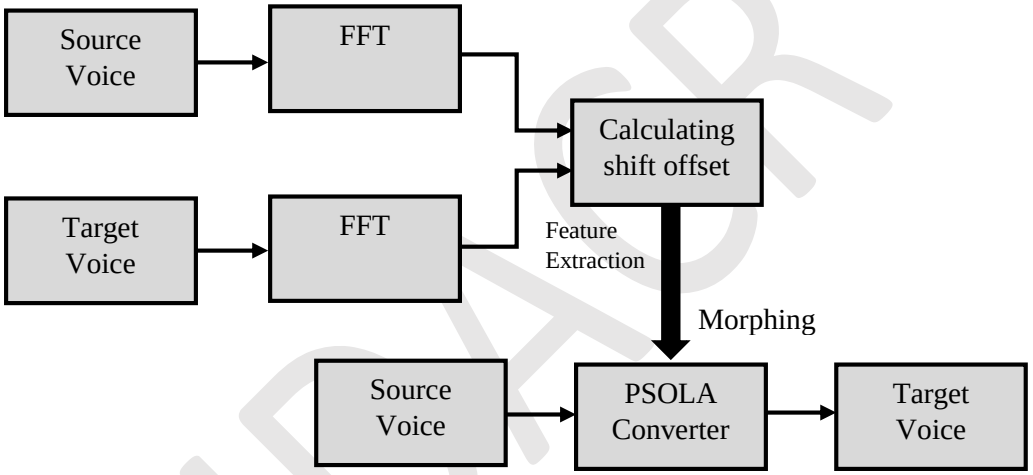

Figure 2: Block diagram for the proposed work

1. Reading source and target files
2. Converting both to frequency domain
3. Finding the difference between fundamental frequencies. Normalized distance in the frequency spectrum will be the shift offset.
4. If shift offset is less than 1 (going for low pitch)
   - Time stretching by shift offset first by SOLA
   - then pitch shifting
5. Else (going for high pitch)
   - Pitch shifting first
   - Then time stretching by shift offset by SOLA
6. Output is the converted file.

Figure 2 shows the block diagram for the proposed work

Pitch Synchronous over Lap-Add (PSOLA) algorithm

One of the approaches in order to change the voice is to change the pitch of the voice; this is done by shifting the pitch of the voice using certain techniques like the Pitch Synchronous over Lap-Add (PSOLA) algorithm.

Pitch Shifting

The pitch period is responsible for making some sounds to be sharper than others. The number of vibrations produced during a given period determines the pitch period. This vibration rate of a sound is named as frequency, higher the frequency higher the pitch. The aim of pitch shifting algorithms is to create a change in pitch without creating a change in the replay rate. Pitch shifting can be done by performing a time stretch using PSOLA and resampling.

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 3, Issue 7, February 2015)

PSOLA

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 [4].

There are several types of PSOLA such as Time Domain TD-PSOLA, Frequency Domain PSOLA (FD-PSOLA) and the Linear-Predictive PSOLA (LP-PSOLA). TDPSOLA is the most commonly used due to its computational efficiency but the others are more appropriate approaches for pitch-scale modifications because they provide independent control over the spectral envelope of the synthesis signal [5].

TD-PSOLA Algorithm

The TD-PSOLA algorithm was proposed allowing pitch modification of a given speech signal without changing the time duration and vice versa [6]. The TD-PSOLA consists mainly of the following three steps:

1. The analysis step, where the original speech signal is first divided into separate but often overlapping short term analysis signals (ST). Short term signals $x_m(n)$ are obtained from the digital speech waveform $x(n)$ by multiplying the signal by a sequence of the pitch synchronous analysis window $hm(n)$ as in Eq. (1):

$$X_m(n) = h_m(t_m - n)x(n) \quad (1)$$

   Here m is an index for the short-time signal.
2. The windows, which are usually Hanning type, are centered on the successive instants, t_m 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. 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 main benefits of time-domain algorithms are:

- Easy to be implemented,
- Give decent result when used on both speech signals given that a small pitch scale factor is used,
- Pitch scaling in the time domain can be through by simply combining time-scaling and sample rate conversion.

III. Simulation and Results

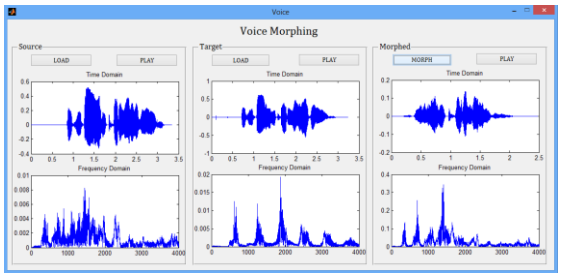

Figure 3: Graphical User Interface (GUI) for proposed work

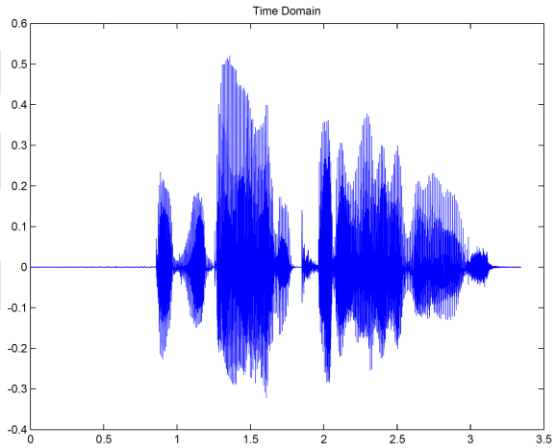

Figure 4: Original source signal

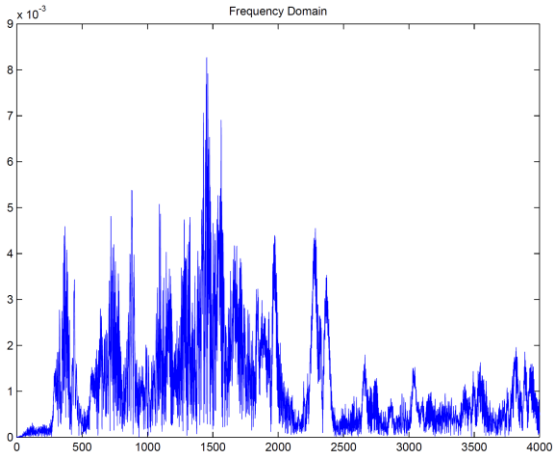

Figure 5: Frequency spectrum of source signal

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 3, Issue 7, February 2015)

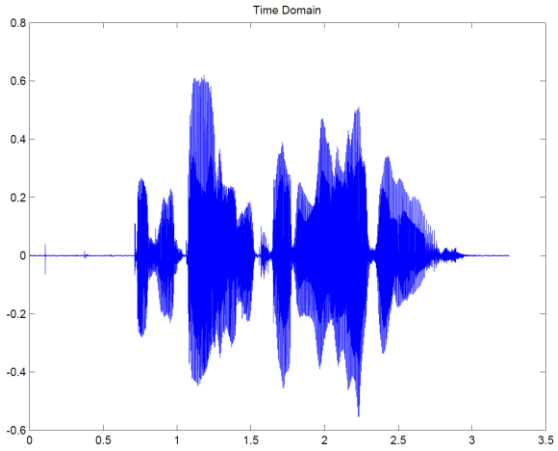

Figure 6: Original target signal

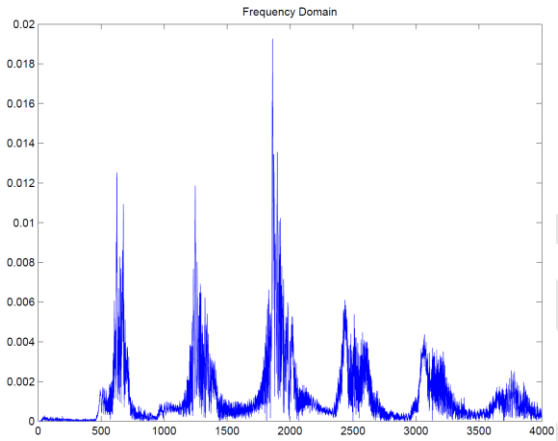

Figure 7: Frequency spectrum of target signal

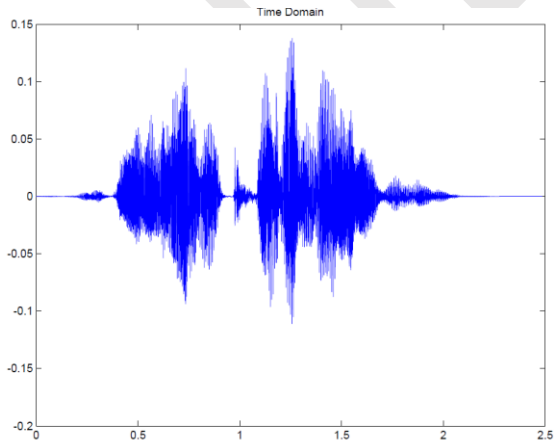

Figure 8: Original morphed signal

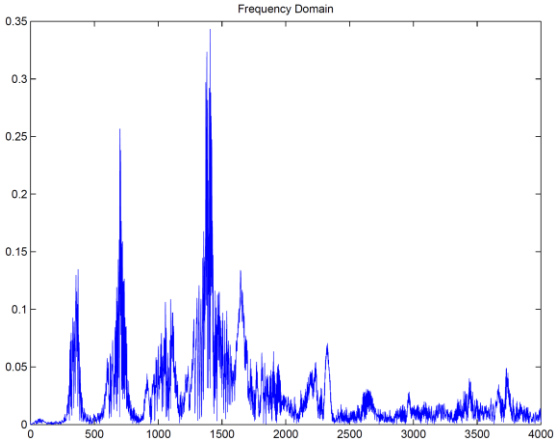

Figure 9: Frequency spectrum of morphed signal

Table 1: Result Comparison

	Source Voice	**Target Voice**	**Morphed Voice**
Fundamental Frequency	437.06	572.98	566.25
Spectrum Power	1.57	1.71	1.71

IV. CONCLUSION

There are basically three interdependent issues that must be solved before building a voice morphing system. Initially, it is vital to improve a mathematical model to signify the speech signal so that the synthetic speech can be redeveloped and prosody can be manipulated deprived of artifacts. Furthermore, the various acoustic cues which enable humans to identify speakers must be identified and extracted. Finally, the type of conversion job and the technique of training must be decided. Here we presented a very simple scheme to produce voice morphing. Pitch of source is converted target using the PSOLA algorithm with time stretching of signal. Results show that, it efficiently convert voice signal from one pitch to another.

REFERENCES

[1] Mireia Farrus, Daniel Erro, and Javier Hernando, "Speaker Recognition Robustness to Voice Conversion", 2008.
[2] Mireia Farrús, Michael Wagner, Daniel Erro and Javier Hernando, "Automatic speaker recognition as a measurement of voice imitation and conversion", 2010.
[3] Helenca Duxans i Barrobes, "Voice Conversion applied to Text-to-Speech systems", Barcelona, May 2006.
[4] Parmod Kumar and Anuj "VHDL Implementation of Phase Vocoder for Voice Morphing", International Journal of New Trends in Electronics and Communication, 2013.
[5] J. H. Nirmal, Suparva Patnaik, and Mukesh A.Zaveri "Line Spectral Pairs Based Voice Conversion using Radial Basis Function", ACEEE Int. J. on Signal & Image Processing, Vol. 4, No. 2, May 2013.

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary research
Website: www.ijdacr.com (Volume 3, Issue 7, February 2015)

[6] Radhika Karthikeyan and G.Ayyappan, "Cepstral approach in voice morphing", Computer Science and Engineering, 2013.
[7] Dharavathu Vijaya Babu, R. V. Krishnaiah "Voice Morphing System for People Suffering from Laryngectomy", International Journal of Science and Research, ISSN: 2319-7064, 2013.
[8] Shaik Shafee, B. Anuradha "Voice Conversion Using Different Pitch Shifting Approach over TD-PSOLA Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, 2013.
[9] Jagannath H Nirmal, Mukesh A Zaveri, Suprava Patnaik and Pramod H Kachare "A novel voice conversion approach using admissible wavelet packet decomposition", EURASIP Journal on Audio, Speech, andMusic Processing 2013.
[10] Yogesh Ganvit, M.A.Lokhandwala, Ninad S. Bhatt "Implementation & Overall Performance Evaluation Of Voice Morphing Based On PSOLA", International Journal of Advanced Engineering Technology, E-ISSN 0976-3945, 2012.