International Journal of Computer Techniques -– Volume 2 Issue2, Mar – Apr 2015

RESEARCH ARTICLE OPEN ACCESS

Voice morphing

Ms.P V Gujarathi[1], Mrs. S S Dinde[2]
[1](E & TC, Pune/ RSCOE Tathawade, Pune)
[2](E & TC, Pune/ RSCOE Tathawade, Pune)

--************************----------------------------------

Abstract:

Voice conversion is a process of transforming the parameters of a source voice to those of a target voice. It is the process of compressing or stretching the time basis of the speech signal without changing its spectral contents. The second is to change the tone of the speech signal, called pitch scaling. It modifies the spectrum of the signal without changing its playback time. These two speech transformation functions are important in many applications such as speech transmission and storage, audio-visual systems, speech recognition, and text to speech conversion. speech modification is done by using pitch scale modification, spectrum modification. Different parameters can be modified like pitch level, pitch range, pitch contour, pitch variation using pitch scale modification. The aim for speech morphing is to produce a smooth change in speech identity.

Keywords **— Spectrum modification**.

--************************----------------------------------

I. INTRODUCTION

A speech signal consist of different frequency which are harmonically related to each other in the form of series. The lowest freq of this harmonic series is known as fundamental freq or pitch freq. Pitch freq is the fundamental freq of vibrations of the vocal cords. This freq generated by vocal cords in the form of periodic excitation passes via vocal tract filter and gets convolved with the impulse response of the filter to produce a speech signal. Thus speech is basically a convolved signal.

The main principle used in time domain pitch detection algorithm is to find similarity between original speech signal and its shifted version.

The main principle used in time domain pitch detection algorithm is to find similarity between original speech signal and its shifted version.

Features of speech segment:
1)Fundamental frequency
2) Formants
3) LPC
4))Mel freq cepstral coefficient

There are two types of features: Time domain feature, Transform domain feature.
Transform domain feature is classified as:
1) Frequency domain
2) Cepstral domain
3) Discrete cosine domain
4) Wavelet domain.

A speech consists of different frequencies which are harmonically related to each other in the form of a series. The lowest frequency of this harmonic series is known as the fundamental frequency or pitch frequency. There are different pitch modification parameters as illustrated below:

Pitch level: The modification in the pitch level means the overall level of F0 contour is shifted by multiplying all pitch valves with a rate factor. (Rate =0 implies no change). When the rate valve is high the pitch value undergoes stronger changes than when the rate is low.

II. EXPLANATION

Inputs of the speech morphing algorithm are speech uttered by speaker A and speaker B, and they are assumed to contain the same phoneme sequences. Output consists of speaker A's speech, modified speech, and speaker B's speech. By temporally controlling speech parameters, the identity of the. modified speech gradual changes

III METHODS FOR SPEECH MODIFICATION

1. PITCH DETECTION USING AUTOCORRELATION METHOD

Autocorrelation is the correlation of signal with itself. It is the similarity between samples as a function of the time separation between them. It can be considered as mathematical tool to find repeating patterns and their periods. Autocorrelation methods needs to list two pitch periods to detect pitch periods to detect pitch. An algorithm to detect autocorrelation can be described as follows:

1. Divide the speech in number of segments. Take speech in number of segments. Take a speech segment which is at least two periods. we will use a speech file having sampling frequency of 8khz.so the sampling interval is 0.0125ms.let us consider speech segment of size 400 samples.
2. We will calculate for say 45 overlapping samples. This means that two segment ,one extending over sample numbers 1-45 is correlated to the other segment from the sample numbers 2-46,so on .hence sample shifts in steps of 1 starting from 1.we will use a shift up to 400 samples to find the shift value for which the correlation is highest. The distance between two successive maxima in correlation will give a pitch period in terms of number of samples.
3. Once pitch period for both speakers are found, pscale is calculated and according to pscale target speech is converted. once pitch modification is done then spectrum is modified.

2. SPECTRUM MODIFICATION

There are different methods for spectrum modification of speech. It can be done using homomorphic coder and using sinusoidal coder.

1.The cepstrum obtained from original signal will contain information regarding the vocal tract in low time region and pitch information in the high time region. The low time region is passed via low time lifter which has information of the vocal tract response. After DFT ,log spectrum of vocal tract which is the envelope of the actual spectrum. Then modify the spectrum and take exponentiation to get modified spectrum of vocal tract. Then IDFT block will convert the spectrum into time domain to give the impulse response of vocal tract which is modified.

IV RESULT

Our main task is to convert male voice in to female or female voice into male. So first database is created with same phoneme uttered by female speaker and male speaker. Once database is created pitch period of both female and male are found. Once pitch periods are calculated pitch mark correspondence are found and pscale is calculated. and pitch is modified.

Modified pitch acts as input to next system implemented for spectrum modification. So cepstrum of pitch modified file and original male file is calculated, depending upon that will get by how much amount modification is required. and finally will get cepstrum modified output.

In this database "**Good Morning**" word is uttered by female speaker and male speaker.

Figure shows original male and female speech uttered by same phoneme.

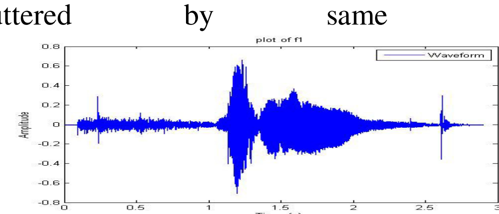

Fig 1 Source Female Speech

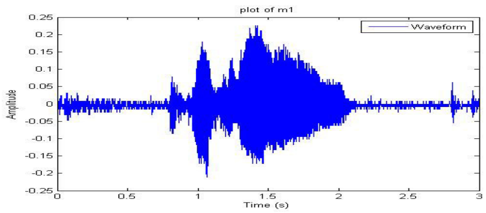

Fig 2 Target Male Speech

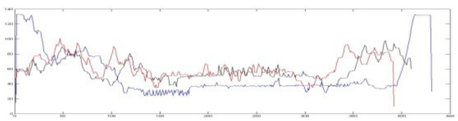

Fig 3 Comparison of male and female and modified pitch period.

TABLE I

Source speaker	Target speaker	Percentage correlation
F1	M1	86.51%
F2	M2	85.70%
F3	M3	65.58%
F4	M4	80.50%
F5	M5	79.80%
F6	M6	86%
F7	M7	86.21%
F8	M8	85.50%
F9	M9	85.23%
F10	M10	82.20%

In this table, percentage correlation between original male voiced and converted voice is given for 10 male and 10 female database

V .CONCLUSIONS

From correlation table it is clear that for most of the voice conversion is done with 85% and subjective listening test average score is 4.14.so high quality results is obtained. So high quality voice conversion system is implemented

REFERENCES

[1] Masanobu Abe, "*Speech morphing by gradually changing spectrum parameter and fundamental frequency*", NTT Human Interface Laboratories.

[2] Masanobu ABE, Shin'ichi TAMURA, Hisao KUWABARA ATR," A new speech modification method by signal reconstruction", Interpreting Telephony Research Laboratories, Japan

[3] C Hamon, E. Moulines, F. Chaarpentier *"A diphone synthesis system based on time domain prosodic modification of speech*", ICASSP'89 pp.238-241,1989.

[4] Sujan Kumar, Md. Kamrul Hasan *"Dominant Harmonic Modification with Data Adaptive Filter Based Algorithm for Robust Pitch Estimation*" Dept. of CSE, University of Rajshahi, Rajshahi, Bangladesh

[5] Daryush Mehta and Thomas F. Quatieri *,"Synthesis , analysis and pitch modification of the breathy vowel*" , 2005 IEEE Workshop on applications of signal processing to audio and acoustics.

[6] E Bryan George member, Mark J. T. Smith *,"Speech analysis / synthesis and modification using an analysis by synthesis/overlap-add sinusoidal model"* ,IEEE transactions on speech and audio processing VOL 5 ,Sep 1997.

[7] Yoshito Nejime, Toshiki Imamura, "*A portable digital speech rate converter for hearing impairment"* , IEEE Transactions on rehabilitation engg, Vol.4, 1996.

[8] Chih-Ting Kuo, Hsiao-Chuan Wang, *"A pitch synchronous method for speech modification*", Department of Electrical Engineering, National Tsing Hua University, Hsinchu.

[9] Vagner L. Latsch and Sergio L. Netto, *"Pitch –Synchronous Time alignment of Speech Signals for Prosody Transplantation.*", Federal University of Rio de Janeiro, Brazil.

[10] Lunji Qiu, Soo-Ngee Koh and Haiyun Yang, "*Pitch determination of Noisy Speech using Wavelet transform in time and frequency domains*", Nanyang Technological University, Singapore.

[11] Gang-Janp Lin, Sau-Gee Chen & Terry Wu," *High Quality **and** Low Complexity Pitch Modification of Acoustic Signals*", Department of Electronics Engineering & Institute of Electronics, National Chiao Tung University Hsinchu, Taiwan, ROC

[12] Michael W. Macon and Mark A. Clements,"*Sinusoidal Modeling and Modification of Unvoiced Speech*", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 5, NO. 6, NOVEMBER 1997 557