Speech Modification by changing Pitch and Spectral parameter

Ms.P.V. Gujarathi, Prof. Dr. Mrs. D Apte.

Abstract— Speech sounds are produced by air pressure vibrations generated by pushing inhaled air from the lungs through the vibrating vocal cords and vocal tract and out from the lips and nose airways. The air is modulated and shaped by the vibrations of the glottal cords, the resonance of the vocal track and nasal cavities, the position of the tongue and the openings and closings of the mouth. Speech transformation changes the speech signals in two directions. The first is to change the speed of the speech signal in the time domain, called time scaling. 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. The time scaling modification is relatively easy that can be performed by TD-PSOLA technique with very good performance. In speech modification there are several methods for time scale modification and different method for spectrum modification of speech. By using speech modification different parameters can be modified like pitch level, pitch range, pitch contour, pitch variation. It can be used to slow down or speed up the speech.

Index Terms— Speech uttered by speakers, pitch modification method (AMDF Method).

1. Introduction

PPitch frequency: A speech signal consist of different freq 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 coards.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.

Features of speech segment:
1) Fundmental frequency
2) Formants
3) LPC

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.

a.Pitch Range: It refers to dynamic range of pitch values the variation in pitch range can be achieved by a shift of each F0-Value by a percentile of its distance to the mean F0-Value of the last syllable range is equal to zero; all pitch values become the last syllable's mean pitch value. If the range value is high the pitch value is shifted from the mean pitch value of the syllable by a large amount, thereby increasing the dynamic range of pitch values.

a. Pitch variation: A pitch variation on the syllable-level is achieved by the application of the pitch range algorithm on each syllable separately. The reference value in this case is the syllable's mean pitch. Using this parameter pitch can be modified.

2. Explanation

Inputs of the speech morphing algorithm axe 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. The control parameters are fundamental frequency (F0) and speech spectrum.

3. Methods for Speech Modification

1. Time Scale Modification /Duration Modification

The speech rate can be modified for the whole phrase. The rate change can be executed by changing the duration of the phonemes. If the duration in consequence of a length reduction is shorter than the frame rate, the phoneme gets dropped. Time scale modification methods are:

a. Consider the problem of 2:1 speed up. Here we need compress a factor of 2. Let the speed signal can be divided in two segments of size 20ms each. To speed it up by a factor of 2, we delete alternate segments of speech. The remaining segments are properly concatenated. To speed up by say 1.5:1, we will have to cut the segment of half the length. The speech will be divided into segment of 20ms and 10ms alternately. And the 10ms segments will be deleted.

b. The first stage is to find the pitch period. The pith signal is decomposed into segment of size equal to one or two pitch periods. The alternate signal will be deleted for 2:1 speed up. In case of 1.5:1 speed up, alternate segments will have length equal to 2. Pitch period and one pitch period. Again the segments of size one pitch period will be deleted it is observed that the speed up is performed pitch synchronously; intelligibility is better than what is obtained. That is without pitch synchronization. Here, the segments must be voiced segments.

2. Spectrum Modification

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

b. 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

ISSN 2229-5518

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 track.Then IDFT block will convert the spectrum into time domain to give the impulse response of vocal track which is modified.

4. Algorithm for Average Magnitude Difference Function

In this we form the difference signal Dm by delaying input speech by various amounts, and subtracting the delayed waveform from the original and summing the magnitude of the differences between sample values. Finally we take the average of the difference function over the number of samples.

1) Divide the speech in the number of segments. First take a speech segment is at least equal to two pitch periods. Let us consider a speech segment of size 400 samples. Then calculate AMDF for say 45 overlapping samples. Hence two speech segment, one is extending over sample numbers 1-45 iscorrelated to the segment extending oversample 2-46,then the segment of sample numbers 3-47that is with a shift of 1,2,3 samples and so on.

Then find the shift value for which AMDF value is smallest. The distance between two successive minima in AMDF will give a pitch period in terms of number of samples.

2) Calculate the AMDF using the formula given by,

AMDF(k)=$\sum_{i=1}^{N}|a(i) - a(i+k)|/N$

3) We have tracked the voice part of speech signal using the fseek command.

5. FIGURES AND TABLES

Table of Pitch Period

Sr Number	Pitch period of female	Pitch period of male
1	90	50
2	160	100
3	260	140
4	330	220
5	360	250
6	400	300

Table of Frequency

Sr No	Frequency of Male	Frequency of Female
1	190	285
2	629	414
3	1073	651
4	1266	985
5	1766	1153
6	2652	1334

6. Results Using Matlab

Fig. 1 Plot of Male Voiced Part

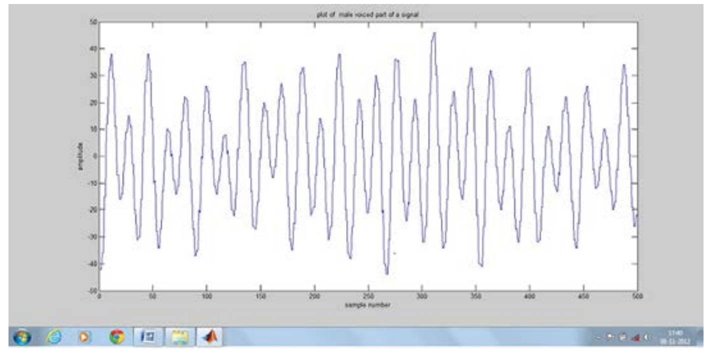


Fig. 2 Plot of Female Voiced Part

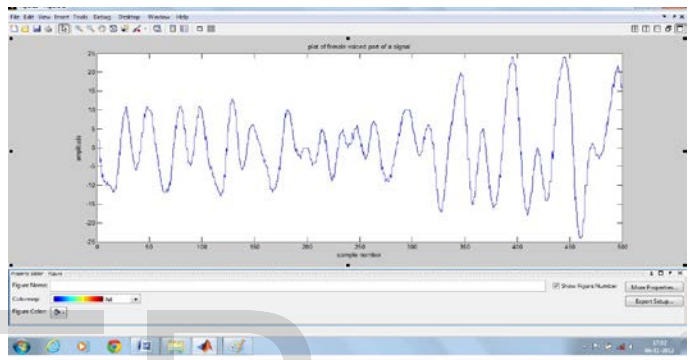


Fig. 3 Plot of AMDF for Male Voiced Part

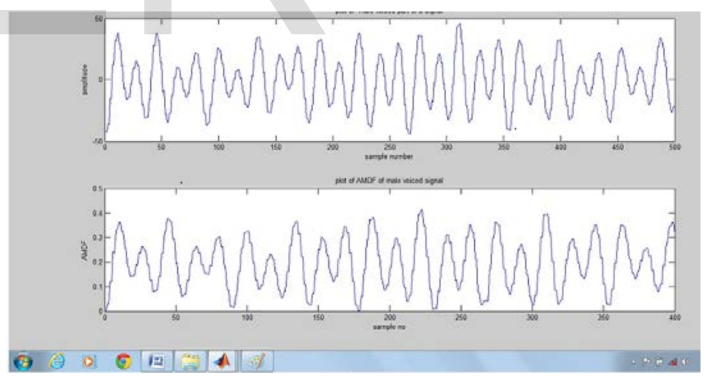


Fig. 4 Plot of AMDF for Female Voiced Part

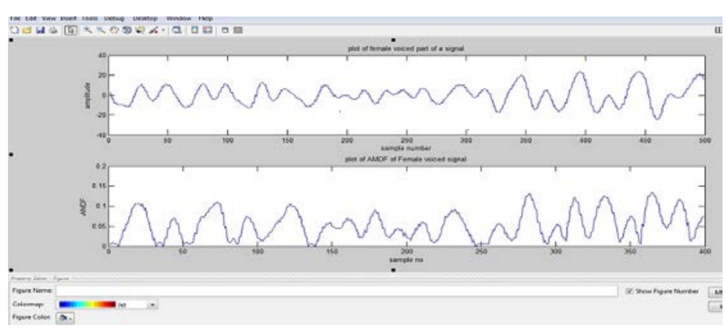


Fig. 5 Plot of Pitch Period Representation of Male Voiced Part

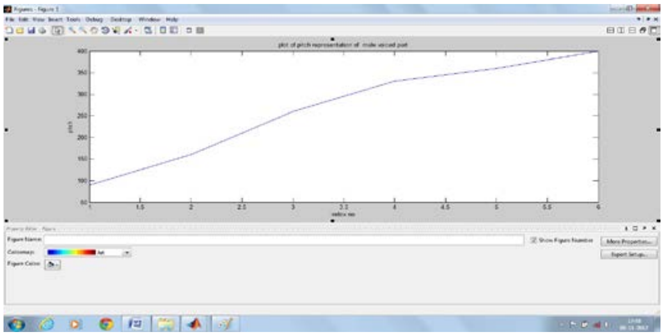


Fig. 6 Plot of Pitch Period Representation of Female Voiced Part

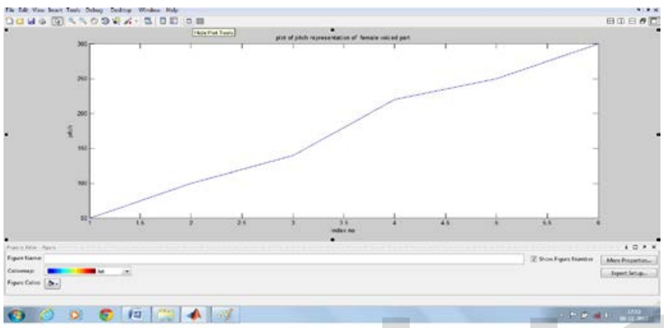


Fig. 7 Plot of Frequency Representation of Female Voiced Part

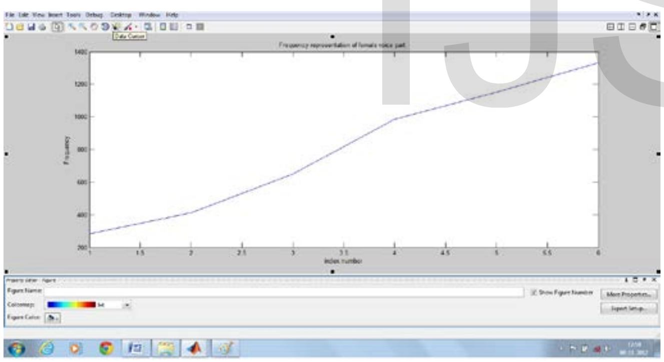
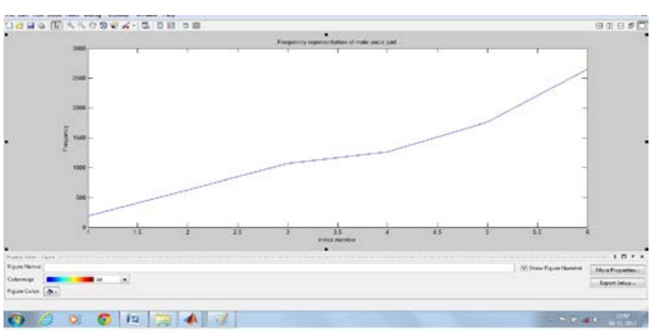


Fig. 8 Plot of Frequency Representation of Male Voiced Part



7. CONCLUSION

A speech signal from a speaker will be recorded .A suitable pitch period method is used i.e. AMDF method to find pitch period. And using pratt software pitch period and spectrum is analyzed. A suitable time duration modification method or pitch modification method and spectrum modification method will be utilized for voice conversion

8. REFERENCES

[1] Masanobu Abe, “Speech morphing by gradually changing spec trum parameter and fundamental frequency”, NTT Human Interface Laboratories.

[2] C Hamon, E. Moulines, F. Chaarpentier “A diphone synthesis system based on time domain prosodic modification of speech”, ICASSP’89 pp.238-241,1989.

[3] 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.

[4] 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.

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

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

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

[8] 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.

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

[10] 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

[11] 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

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

[13] Thomas Baran, Nicolas Malyska, Thomas F “Preserving The Character Of Perturbations In Scaled Pitch Contours”, Quatieri MIT Lincoln Laboratory

[14] Dima Ruinskiy and Yizhar Lavner ,“Stochastic Models Of Pitch Jitter And Amplitude Shimmer For Voice Modification”, Department of Computer Science, Tel-Hai Academic College, Israel

ISSN 2229-5518

[15] R. Muralishankar, M. Ravi Shanker and A. G. Ramakrishnan PES Centre for Intelligent Systems," Perceptual-MVDR Based Analysis-Synthesis Of Pitch Synchronous Frames For Pitch Modification", Department of Telecommunication Engineering, PESIT, Bangalore, INDI

Acknowledgment

I wish to thank Dr Prof Mrs S.D.Apte for her valuable guidance.