RESEARCH ARTICLE OPEN ACCESS

A Novel Approach to Reduce Average Spectrum Distortion in Pitch Alteration Technique Using Gaussian Method

Anirudh Boddapati*, V Ratna Kumari**
*Research Scholar, Department of Electronics & Communication Engineering, PVPSIT, Vijayawada
** Assistant Professor, Department of Electronics & Communication Engineering, PVPSIT, Vijayawada

ABSTRACT
In case of speech synthesis, Intonation plays an important role. Intonation is the variation of pitch. It is used for a range of functions such as indicating the attitudes and emotions of the speaker, signaling the difference between statements and questions, and between different types of question, focusing attention on important elements of the spoken message and also helping to regulate conversational interaction. To alter the pitch, Pitch Synchronous Overlap and Add technique (PSOLA) can be used. But, in the PSOLA technique, applying symmetric window function to asymmetric speech waveform causes energy unbalance phenomenon. This energy unbalance can be overcome by the time-frequency conversion of asymmetric waveform to symmetric. But this conventional method leads to more spectrum distortion. So, we propose a new Gaussian method which can reduce the spectrum distortion over conventional method.
Keywords: Pitch Synchronous Overlap and Add (PSOLA), Gaussian method

I. INTRODUCTION

Text to speech synthesis is used in many applications. For example, it is used as an assistive technology for people with various disabilities. For people with vision impairment [1],a text to speech synthesizer can be used to read email and web pages and other electronic text material. A synthesizer can also aid people with severe speech impairment by a voice-output communication aid, which is a device that produces synthetic speech. A speech synthesizer is also commonly used by people with dyslexia to read or to check self-written text by listening. Another application is human machine speech communication when a large vocabulary is needed or desirable[2].

A method is proposed by Jong-soon Jung, Jeong-jin Kim and Myung-jin Bae, which explains speech reconstruction using waveform symmetry and the PSOLA method to solve the energy unbalance caused by pitch alteration. Because the PSOLA method is already proposed, this method changes asymmetric waveform to symmetric waveform in the time domain to be appropriate to a new PSOLA[3]. But the spectrum distortion is raised in this method. So we can apply Gaussian pitch filter to the above method to reduce average spectrum distortion.

II. Conventional method

Since PSOLA method processes a speech signal in the time domain , it improves the error of modeling of speech production and the spectrum distortion. In addition, it is appropriate to real-time prosody control because it consumes a fewer time due to processing in the time domain. It causes energy unbalance since adapting a symmetric window to an asymmetric speech signal.

2.1 Analysis

If a speech signal is voiced sound, the speech signal is made into a train of short-term signal after multiplying a window function by a decomposed pitch period. If the speech is unvoiced sound, it is analyzed with 10ms .Hanning window is used in the analysis
Hanning Window :

$$W(n)=\frac{1}{2}\left\{1-\cos\left(2\pi\frac{n}{N-1}\right)\right\},0\le n\le N-1$$

Decomposed pitch period is obtained like below Equation by multiplying a speech signal by window function which has symmetric property.

$$S_{analysis}(n)=W_{analysis}(m-n)S(n)$$

$S_{anlysis}(n)$: ST(Short Term) signal of pitch period

$W_{analysis}(n)$: analysis window function

m:m_th pitch

S(n):original speech signal

2.2 Prosody control and synthesis

The train of ST signal is arranged by pitch period of the speech signal. Therefore, in order to change the pitch, pitch period is rearranged by altered pitch period. Following Equation represents a pitch altered signal.

$$S_{synthesis}(n)=S_{analysis}(n-m_a)$$

$S_{synthesis}(n)$: Pitch altered ST signal

m_a : Pith period to be altered

Therefore when we want pitch to make high, the period of ST signal is arranged small and when we want pitch to make low, the period of ST signal is arranged large. However, it is important to maintain accurate pitch synchronization between sequential arrangements. Also, overlapped part of a rearranged ST signal is simply added.

The PSOLA synthesis technique makes an original speech signal in to a train of ST signal train by multiplying decomposed pitch period with window function. Speech is synthesized from a controlled unit after the prosody control. However, the PSOLA synthesis technique applies symmetric window even though speech signal is asymmetric. Therefore, energy unbalance is caused according to the degree of overlap when the pitch period is controlled. Thus, normalization in order to make energy constant is required. So asymmetric speech signal is changed into symmetric signal using energy normalization in the time domain conversion, which is appropriate to the PSOLA method[3].

But spectrum distortion is increased in this method. So we propose a new Gaussian method to reduce the spectrum distortion

III. Proposed Gaussian Method

For a D-dimensional input vector o, the Gaussian distribution with mean μ and a positive definite covariance matrix $\sum$ can be expressed as

$$N(o,\mu,\sum)=(2\pi)^{-\frac{D}{2}}(\sum)^{-\frac{1}{2}}e^{-\frac{1}{2}(o-\mu)^T\sum^{-1}(o-\mu)}$$

3.1 Gaussian filter:

Gaussian filter is a linear filter which is used in pitch alteration applications in speech processing. It minimizes the rise and fall time. So it is used as a speech smoother. The Gaussian filter is non-causal which means the filter window is speech symmetric about the origin. This gives an output speech pulse shaped like a Gaussian function. Mathematically, a speech Gaussian filter modifies the input signal by convolution with a Gaussian function; this transformation is also known as the Weierstrass transform. We will apply the Gaussian filter to the conventional system to reduce spectrum distortion[4].

3.2 Edge Based Region(EBR) Filter:

Edge based Region filter is used in the field of voice deduction and Speech analysis. Specifically the EBR detector is designed for speech compression and expansion. EBR filters can typically be classified into two categories: intensity-based detectors and structure-based detectors.

3.2.1 Intensity-based detectors

It depends on analyzing local differential geometry or intensity patterns to find points or regions that satisfy some uniqueness and stability criteria. These detectors include Scale-invariant Feature Transform(SIFT) etc.

3.2.2 Structure-based detectors

It depends on structural features such as lines, edges and curves, etc. to define the interest points or regions. These detectors include scale-invariant shape features (SISF) etc.

3.3 Munich and Cambridge Morlet wavelet Filters

Morlet wavelet is chosen as the mother wavelet. We use the Morlet wavelet based on the fact that it has good properties in joint time frequency localization and has a well defined impulse response. In the time domain, the mother wavelet is a high-frequency vibration whose amplitude decreases when the time varies from zero to infinity. The Morlet wavelet transform has proved to be beneficial for many types of wave signal processing and has shown a good performance in tasks like speech compression ,expansion and audio coding [5].

IV. SIMULATION RESULTS

Four sentences which are given as input are given below. They are sampled at 15 kHz and quantized at 16 bits

Sentences 1: Hello ! I am Anirudh.

Sentences2: The Lecture Takes Place On Thu/10/10 2013. The Lecture Starts at 1 Pm.

Sentences3: He Made A Record and I Will Record the time.

Sentences4: May I help you ?

The above signal is separated based on pitch period and symmetry is applied to voiced sound only. To alter the pitch, compression and expansion of the signal is performed. Table 1 and Table 2 show the objective evaluation of pitch compression and expansion. Table 3 shows the comparison between conventional method and proposed method. Proposed method is better than the conventional method as it reduces the average distortion ratio by 2%.

Table1. Spectrum distortion of pitch compression

Compression Ratio(%)	90	80	70	60	50
Distortion(%)	4.9	6.5	8.2	11.5	14.6

Table2.Spectrum distortion of pitch expansion

Expansion Ratio(%)	120	140	160	180	200
Distortion(%)	2.34	3.2	4.18	7.9	11.3

Table3. Comparison of compression and expansion ratios between the conventional method and proposed method

Method	Compression Ratio	Expansion Ratio
Conventional	101.67	201.67
Proposed	73.66	145.06

Table 4.Comparison of average spectrum distortion between conventional method and proposed method

Method	Average Spectrum Distortion Ratio
Conventional	7.46
Proposed	5.65

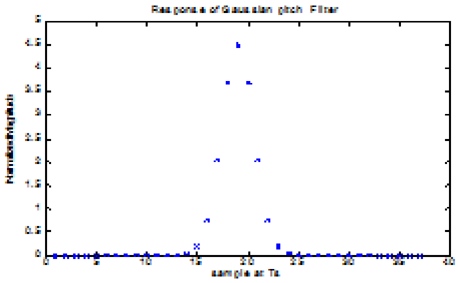

Figure 1 Response of Gaussian pitch filter

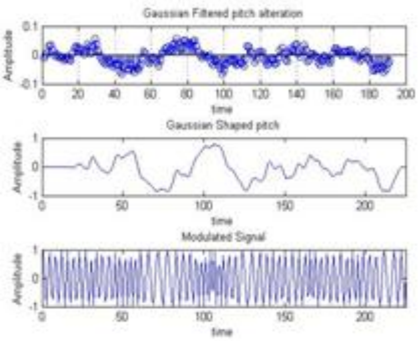

Figure 2 Pitch alterations for Gaussian filter and modulated signal

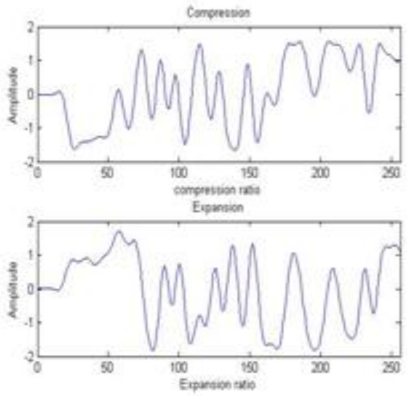

Figure 3 compression and expansion of the input signal

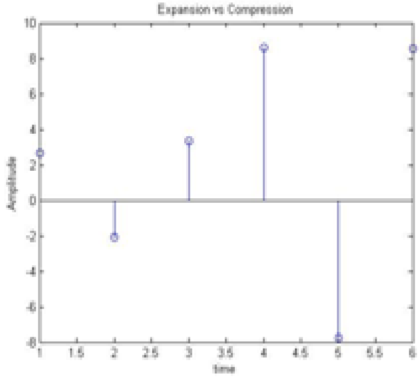

Figure 4 Comparison of compression and expansion

Figure (1) indicates the response of Gaussian filter for given signal. Figure (2) indicates the Gaussian pitch alterations and modulated signal. Figure (3) indicates the compression and expansion of the input signal. Figure (4) indicates the comparison between compression and expansion.

V. CONCLUSION

For pitch alteration, Pitch Synchronous Overlap and Add technique can be used. But, in PSOLA applying symmetric window to a asymmetric signal causes energy unbalance phenomenon. A Pitch Alteration method is proposed which can solve the energy unbalance. This method separates the waveform or pitch period in order to solve the energy unbalance caused by the degree of the overlap by controlling the pitch period. In this conventional method spectrum distortion is higher. Our proposed Gaussian method reduces the spectrum distortion by 2%.

REFERENCES

[1] I E.MOULINES, F. CHARPENTIER, "Pitch Synchronous waveform Processing techniques for Text-To-Speech Synthesis using diphones", Speech Communication, Vol. 9, n 5-6. 1989.

[2] L.R Rabiner & R.W. Schafer, Digital Processing of Speech Singals, Prentice-Hall, 1978.

[3] Jong-soon Jung , Jeong-jin Kim and Myung-jin Bae, "Pitch Alteration Technique In Speech Synthesis System," IEEE, Transactions on Consumer Electronics, Vol. 47, No. 1, February 2001.
[4] Shah A., Ramachandran, R.P. ; Lewis, M.A. "Robust pitch estimation using an event based adaptive Gaussian derivative filter " Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on (Volume:2).
[5] S. China Venkateswarlu, V. Sridhar, A. Subba Rami Reddy & K. Satya Prasad "Audio Compression using Munich and Cambridge Filters for Audio Coding with Morlet Wavelet" Global Journal of Computer Science and Technology Software & Data Engineering Volume 13 Issue 5 Version 1.0 Year 2013