

The Comparison of the Effect of Haimming Window and Blackman Window in the Time-Scaling and Pitch-Shifting Algorithms

LIN Zhiwei², DA Li², WANG Hao¹, HAN Wei¹, LIN Fan¹

1. Software school, Xiamen University, Xiamen, 361005, China,

2. Department of Computer Scientist, Xiamen University, Xiamen, 361005, China

iamafan@xmu.edu.cn

Keywords: pitch shifting Haimming Window Blackman Window .

Abstract. The real-time pitch shifting process is widely used in various types of music production. The pitch shifting technology can be divided into two major types, the time domain type and the frequency domain type. Compared with the time domain method, the frequency domain method has the advantage of large shifting scale, low total cost of computing and the more flexibility of the algorithm. However, the use of Fourier Transform in frequency domain processing leads to the inevitable inherent frequency leakage effects which decrease the accuracy of the pitch shifting effect. In order to restrain the side effect of Fourier Transform, window functions are used to fall down the spectrum-aliasing. In practical processing, Haimming Window and Blackman Window are frequently used. In this paper, we compare both the effect of the two window functions in the restraint of frequency leakage and the performance and accuracy in subjective based on the traditional phase vocoder[1]. Experiment shows that Haimming Window is generally better than Blackman Window in pitch shifting process.

Introduction

In the point view of frequency, audio can be seen as a discrete signal which composed by a sine wave that changes time by time. Music signal can be seen as a smooth signal in a short period of time (usually 10 ~ 30 ms). It is relatively stable and simple during the period of time. And the voice is a monotonous voice in subjective. Because of this stable feature of music, Short-Time Fourier transformation (STFT)[2] is widely used. This signal is called a frame of the period of time in usual. It can intercept all the frames of time by windowing moved method.

2. Time/Frequency changed Algorithm

Changed the pitch of the audio, is to change the frequency which composes the audio-wave. Pitch changed algorithm is based on this thinking. Double wave's frequency, it will increase G-8 degrees of pitch music in theory. However, pure pitch changed in frequency domain makes each phase inconsistent. Also it makes an echo effect. In addition, in theory, because of the limited of the length of window, the frequency can not be strictly separated, namely frequency spectrum-aliasing. This reason makes us can not analysis the composition of the wave in wave spectrum accurately. Frequency leak and frequency spectrum-aliasing are two statements for one issue. This paper does not distinguish above two statements.

Short-time Fourier transform (SFFT) analysis comprehensive method is an effective solution for solving phase discontinuous. This method makes use of windowing increment, Fourier transform, frequency/phase adaptation, comprehensive windowing and stacking process[3]. Eliminate echo effect effectively, known as phase synthesis.

2.1 Improvement the algorithm of phase synthesis

The traditional phase compose model includes four processes. There are Fourier transform with windowing signal, frequency and phase adaptation, comprehensive windowing and output stacking [4]. In this paper, we use Hamming window as a discrete Fourier transform the operator. After

compare different audio frequency samples. Found that effect is better than Blackman window. The experiment also shows that the Hamming window on the frequency of leakage suppression has a better effect. The sequence after windowing needs reconstruction to restore its original energy. For Blackman window and Haiming window, the reconstruction process can use the same windowing function. It can restore the time-domain signal by using integrated stacking method.

Blackman Window

$$w_h(n) = \begin{cases} 0.42 - 0.5 \cos \frac{2\pi n}{M} + 0.08 \cos \frac{4\pi n}{M} \\ 0 \end{cases} \quad (1)$$

Haiming Window

$$w_m(n) = \begin{cases} 0.54 - 0.46 \cos \frac{2\pi n}{M} \\ 0 \end{cases} \quad (2)$$

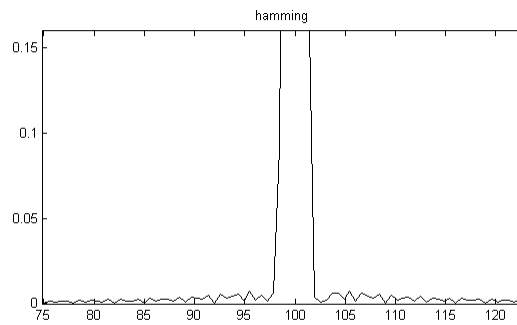


Fig. 1. The spectrum of audio signal with Haimming window function and Fourier transition

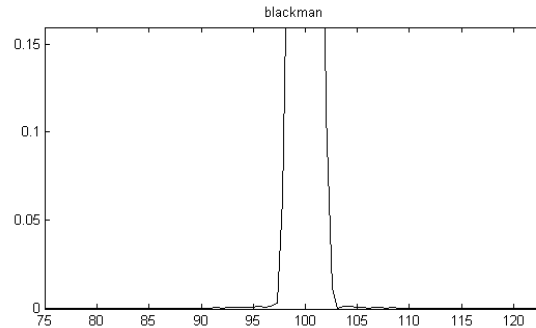


Fig. 2. The spectrum of audio signal with Blackman window function and Fourier transition

Through comparison, Haimming window restrain the signal frequency at a short wide area near 100 Hz. Signal that applied with Haimming window has the feature of narrower main bean, concentration of energy and the more accurate frequency which will improve the performance the pitch shifting process. Blackman widow obviously has a wider main bean with energy leakage. Meanwhile, it should be pointed out that signal with Haimming Window has more side lobes in the experiment which to a certain extent offsets the concentration effect of the narrow main lobe. In general, Haimming Window is better than Blackman Window in the restraint of frequency leakage.

2.2 Time domain windowing and its restoration

Discrete Fourier Transform with Hamming window

$$x_w(n) = x(n) \cdot h(sR - n) \quad (3)$$

Following formulary is the sequence of window reconstruction. Here $f(n)$ stands for reconstruction window

$$x(n) = \frac{1}{M} \sum_{s=-\infty}^{\infty} x_w(n) f(sR - n) \quad (4)$$

2.3 Restoration the pitch changed energy

After pitch changed processing, the scale of time domain window is changed too. In order to restore energy accurately, formulary (5) modifies the reconstruction window.

$$\tilde{x}(n) = \frac{1}{M} \sum_{s=-\infty}^{\infty} x(n) \cdot h_p(sR - n) f_p(sR - n) \quad (5)$$

Here h_p is the pitch changed coefficient. P is the yardstick factorial window. If use w_m to replace h_p , then it can produce the coefficient of Hamming reconstruction window. f_p Could be constructed by a ratio multiplies f . It is the same as restructure a reconstruction coefficient.

$$\begin{aligned} c_p &= \frac{M}{\sum_{s=-\infty}^{\infty} w_m(sR - n) w_m[p \cdot (sR - n)]} \\ &= \frac{M}{\sum_{s=-\infty}^{\infty} (0.54 - 0.46 \cos \frac{2\pi p(sR - n)}{M}) (0.54 - 0.46 \cos \frac{2\pi (sR - n)}{M})} \\ &= \begin{cases} \frac{R}{(0.08 + 0.92^2 \cdot 1/4)} & p \neq 1 \\ \frac{R}{(0.08 + 0.92^2 \cdot 3/8)} & p = 1 \end{cases} \end{aligned} \quad (6)$$

In actual processing, because the sequence's status may be changed after being processed, and the process of comprehensive windowing and stacking can not guarantee restore all of the energy. But in specific circumstances [5], this process can restore most of the energy. The experiments show that Haiming window is better that Blackman window on restraining frequency leakage. In improved algorithms, we use Haiming window as the convolution window in analyzing and integrated process.

2.4 Process of analysis/synthesis

We can use a FIFO queue to receive audio input sequence. The length of enter queue is equivalent to the length of the window. The enter queue processes forward R samples after analyzing every time. In order to restrain the Frequency spectrum-aliasing, it needs to windowing the sequence of enter queue. The window sequence is changed to be a new sequence $\tilde{x}_w(n)$ after being processed. And the new sequence will be stacked up into the last shift output buffer after windowing again to restore original signal energy. The reconstruct signal is $\tilde{x}(n)$. Phase Vocoder [6] algorithm uses the constant 4 as the reconstruct coefficient. The signal energies are different between before and after pitch changed. It does not have a good adaptation for using a constant value. Improved algorithm introduce into a pitch factor for correction factor. It corrects the changed energy in the sense of listening.

2.5 Fast Fourier Transform in audio processing

Fourier transform in audio processing is the most common form of transformation. Transform changes the signal from the time domain to frequency domain, Fourier inverse transform changes the signal from the frequency domain to time domain. In the process of audio signal with computer, it is impossible to measure and compute the signal of an infinite length. The proper method is to cut out a time frame and then apply the periodic continuation method to get a virtue infinite signal before

Fourier transform. The truncated signal spreads its energy by aliasing effect, which is also called frequency leakage. The use of Haimming Window as the operator of Fourier transform has better performance than Blackman window to minimize frequency leakage. During this procedure, Digital audio is the sample result which is a discrete data for analog signals. So we usually said Fourier transform in audio processing referring to the Discrete Fourier Transform (DFT). And its inverse formularies as follows: (7) Fourier Transform. (8) Inverse Fourier Transform.

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/N} \quad (7)$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi nk/N} \quad (k = 0, 1, \dots, N-1) \quad (8)$$

2.6 Process of frequency/phase adaptation

This is the core of the pitch changed. The sequence after Fourier transform called complex sequence. Mapping to the complex plane is the Cartesian coordinate. Pitch changed base on frequency changed and phase modulation, it is necessary that the complex sequence should be denoted as module/phase polar coordinate sequence [7].

$$\begin{aligned} mag &= \sqrt{real^2 + imag^2} \\ phase &= \arctan \frac{imag}{real} \end{aligned} \quad (9)$$

The purpose of transform is to calculate the diff of phase of two samples' analysis sequence of the spectrum components. These two samples are discrepancy R samples. Multiply this phase diff with the pitch changed coefficient to generate a new phase. Restructure frequency complex sequence with this new phase [8]. And map this new sequence to the comprehensive spectrum buffer.

2.7 The relationship of subjective characteristics and the length of window M and number R of skip samples

The larger of M, the larger of scale of the window covers. And the smaller of change error of spectral analyze [9]. Experiment shows that, human ear is sensitive on frequency error in high pitch of music. A large window can increase the accuracy of spectral analysis. For 44.1 KHz audio, let $M > 2048$ can get the pitch changed coefficient between 1 and 2 which bring a good effect. Audio music keep original pitch unchanged when pitch coefficient is equivalent 1.

2.8 Promotion of frequency modulation algorithm

Audio's pitch changed and time stretching (pitch unchanged) can be regarded as one issue [10]. A section of audio which is double sampled, the pitch will be improved eight degree when play this audio, and vice versa. To ensure time changed without pitch changed, or pitch changed without time changed, both must modify the original audio [11]. A simple example is that, play a music which has been improved eight degree with a half sampled rate speed. The pitch is the same as the original one, but the time is double. It can get that same pitch when change audio's pitch using phase integrating algorithm and then linearity interpolate or sample. But the result of new time scale divides original audio frequency is equal to the original pitch changed factor's audio. Experiments show that it also can achieve satisfactory results by phase synthesis [12].

Summary

Part of low frequency of 2 times wave is missed in Phase Vocoder [4] pitch changed algorithm. Improved algorithm manifests the low frequency of original wave and composes a new wave which doubles the original wave in naturally. Divide music into the background instrument and voice. For pitch rising, there is an echo effect of instruments reverberation using frequency modulation method directly to background instruments. Can be heard sound duplication obviously for human voice; for pitch falling, it produces noise when using frequency modulation to the area which between two frame's connection. It solves these two issues effectively by phase composed pitch changed algorithm. The smoothness of voice is low by Phase Vocoder[4]. Improve algorithm keeps the loudness of music consistent between before and after changed of the music in subjective. The quality of voice is improved markedly.

Because of the inherent frequency leakage effects in short term Fourier transform, phase synthesis method makes use of windowing at time domain to fall down the spectrum-aliasing. It makes each of phases inconsistent of analysis frame by directly frequency modulate. It gets a wonderful smooth effect in subjective when connects different phases by adjusting the phases' odds and phases' sum in the processing of phase changed. It changes the original sequence's energy because of window. But it restores the energy by composite window/stacking process once again. Choose appropriate parameters for different factors. Choose large length of window for frequency rises and choose small step rate for frequency falling.

References

- [1] J. Laroche, "Time and pitch scale modification of audio signals," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Kluwer, Norwell, MA, 1998.
- [2] J.L. Flanagan and R.M. Golden, "Phase vocoder," *Bell Syst. Tech. J.*, vol. 45, pp. 1493–1509, Nov 1966.
- [3] J. B. Allen and L. R. rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977.
- [4] R. Portnoff, "Time-scale modifications of speech based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, no. 3, pp. 374–390, 1981.
- [5] M.S. Puckette, "Phase-locked vocoder," in *Proc. IEEE ASSPWorkshop on app. of sig. proc. to audio and acous.*, New Paltz, NY, 1995.
- [6] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," to appear in May issue of *IEEE trans. speech and audio proc.*, 1999.
- [7] L.B. Almeida and F.M. Silva, "Variable-frequency synthesis: an improved harmonic coding scheme," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal processing*, 1984, pp. 27.5.1–27.5.4.
- [8] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 4, pp. 744–754, Aug 1986.
- [9] X. Serra and J. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music J.*, vol. 14, no. 4, pp. 12–24, Winter 1990.
- [10] E. B. George and M. J. T. Smith, "Analysis-bysynthesis/ Overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones," *J. Audio Eng. Soc.*, vol. 40, no. 6, pp. 497–516, 1992.
- [11] S. Tassart and P. Depalle, "Analytical approximations of fractional delays: Lagrange interpolators and allpass filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Munich, Germany, 1997.
- [12] T.I. Laakso, V. Valimaki, M. Karjalainen, and U. KLaine, "Splitting the unit delay [fir/all pass filters design]," *IEEE Signal Processing mag.*, vol. 13, no.1, pp. 30–60, Jan 1996.