The comparison of window functions for different subbands in Phase Vocoder

Xiao Bin [1] Jiang Yi[2].

Abstract. The phase vocoder[1] has long been a well established tool in pitch shifting technology. It analyzes and synthesizes audio signal via short-time Fourier Transform. Unfortunately, Fourier Transform leads to the inevitable inherent frequency leakage which decreases the accuracy of the pitch shifting effect [2]. In order to restrain this side effect, window functions are used to intercept audio. Hamming Window, Hanning Window, Blackman Window and Kaiser Window are well known windows used in digital technology [3][4]. This paper compares the effect of the four windows in restraint of frequency leakage and maintaining the audio quality after pitch shifting. Based on the MP3 encoding principle, we divide the human hearing frequency range 20Hz-20K Hz [5] into 32 equal width subbands. Then, for each sub-band, four window functions are separately used to intercept sound [6]. Moreover, the SVM method is introduced to classify suitable and unsuitable window for each subband. Finally, we conclude the suitable frequency scale for each of the four window functions.

Keywords: Phase vocoder; windowing ; MP3; SVM.

1.1 INTRODUCTION

The real-time pitch shifting process is mainly 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, high degree of flexibility and can be used with other audio processing at the same time. Phase vocoder [7] is the major frequency domain method which shifts the audio pitch by changing the frequency spectrum.

In the frequency view, audio can be seen as a discrete signal composed by many sinusoidal components whose frequencies and amplitudes change over time. As the human hearing frequency is between 20 Hz and 20 kHz, the sine waves frequencies are within this

[1] Xiao Bin (✉)
Department of Computer Science and Technology,Xia Men University,Xiamen,China
e-mail: xiaobin1990@foxmail.com

[2] Jiang Yi
Department of Computer Science and Technology,Xia Men University,Xiamen,China
e-mail: jiangyi@xmu.edu.cn

range. Changing the audio pitch is to change the wave frequency which consists the audio. Pitch shifting algorithm is based on such method that increasing or decreasing twice the wave frequency makes the audio pitch increase or decrease by so called an octave in music theory.

In the process of audio signal with computer, it is impossible to measure and compute the signal of an infinite length. Music signal can be seen as a smooth signal in a short period of time (usually 10 ~ 30 ms). Because of this stable feature of music, Short-Time Fourier transformation (STFT) [8] is widely used to intercept audio into small pieces of signal. This signal is called a frame of the period of time in usual. STFT can intercept all the frames of time by windowing moved method. Then apply the periodic continuation method to get a virtual infinite signal. Unfortunately, in the STFT procedure, the truncated signal spreads its energy to adjacent spectrum. This phenomenon is called frequency leakage [9].So,how to choose the most suitable window become important.

1.2 EXPERIMENT

1.2.1 Windowing Graphic Analysis

With the help of mp3 encoding principle, human hearing frequency range is separated into 32 equivalent width subbands. Each subband will be operated for windowing analysis.We choose three representative subbands separately in low, middle and high frequency for analysis.

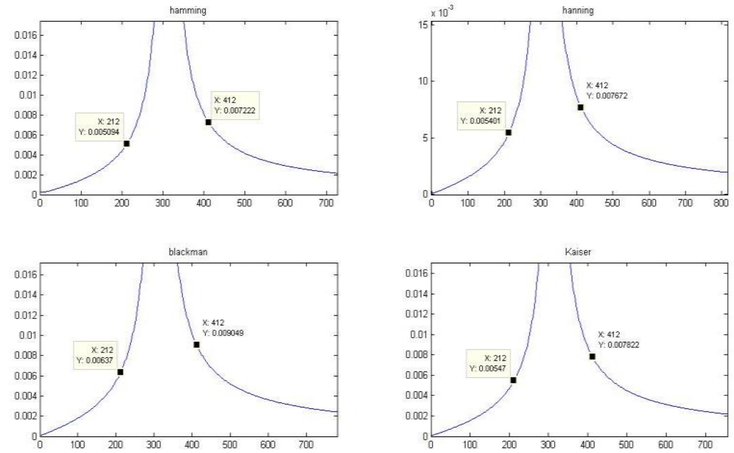

Fig. 1.1 Low frequency subband windowing analysis

The experiment takes the first subband whose central frequency is 312.5Hz as the sample of low frequency subbands. This subband contains fundamental frequency and low overtones frequency. It makes great contribution to fullness of sound. Through pitch shifting process with four window functions, we found audio output processed with Hamming Window remains the highest main-lobe energy. The output can precisely describe the frequency that transformed from time domain to frequency domain. Blackman window has the sharpest slope in the comparison, which shows its fast convergence speed and good performance in restriction of frequency leakage.

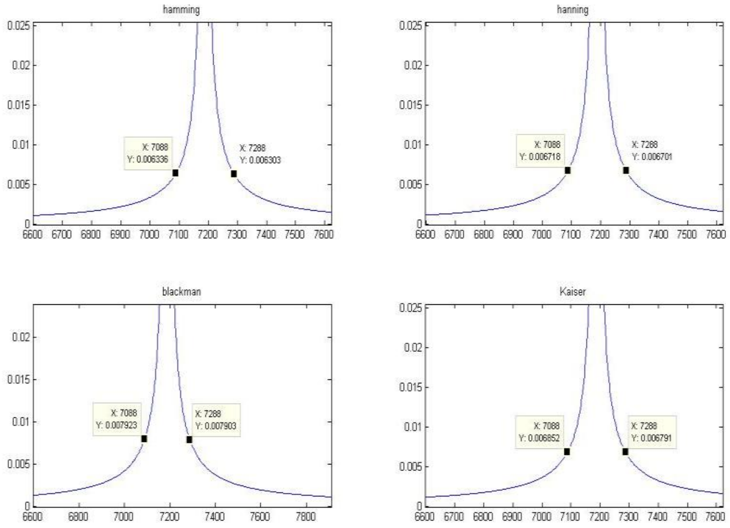

Fig. 1.2 Middle frequency subband windowing analysis

This experiment takes the 12th subband whose central frequency is 7187.5 Hz as the sample of middle frequency subbands. Based on the acoustic sensory pleasantness model, human hearing is sensitive frequency around 7K Hz. Meanwhile, this band contains middle and high frequency overtones, making great contribution to expressive performance of stringed instruments. Therefore, this band has the highest requirement for audio quality. Through the comparison and analysis, Hanning window has the highest slope, namely the fastest convergence speed and a comparatively concentrated M value. Besides, Blackman window also has a good performance in this subband. But Kaiser window spreads wave power to adjacent area obviously, indicating a bad behavior.

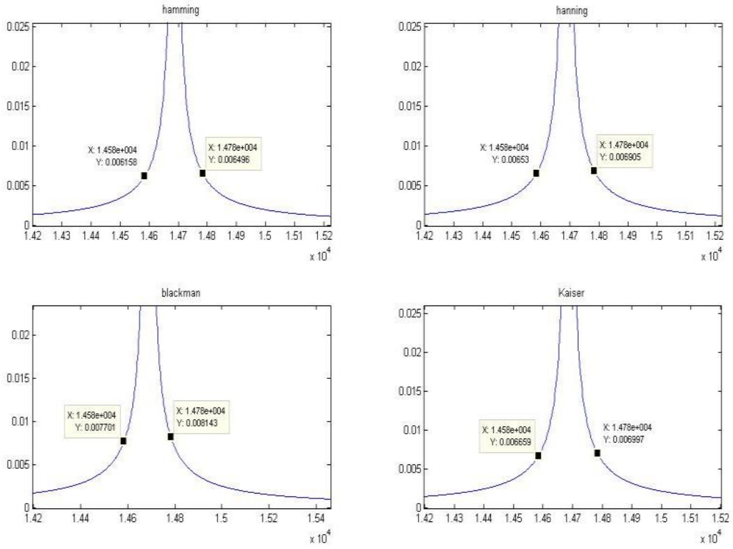

Fig.1. 3 High frequency subband windowing analysis

The experiment takes the 24th subband whose central frequency is14687.5Hz as the sample of middle frequency subbands. This band is not far from the upper limits of human hearing frequency. Acoustic sensory pleasantness model tell us audio of this band should be decayed for proper degree in order to reduce the general sharpness. The experiment results show Hamming window undermines general smoothness and causes discontinuous of adjacent subbands in splicing process. Therefore Hamming window is not recommended for Fourier transform in high frequency subbands. On the other hand, other three windows could improve the transition process smoothly and satisfies the requirement for the high frequency subbands.

1.2.2 Classification of Window function based on SVM

1.2.2.1 Music feature abstraction:

In order to evaluate the output music after pitch shifting with quantizing analysis method， we define a serious music features. The experiment abstracts feature values from the music that has been processed with window functions.

The music features to be abstracted are as follow. Here, $S_n(i)$ stands for sampling signal from No. n frame, length N; $F_n(j)$ is the sampling sequence of frequency, length M; $W_n\ (j)$ stands for related value in spectrum of $F_n(j)$, namely $W_n^{'}(F_n(j))$.

- Overall intensity

Intensity is the loudness of music. From physical aspect, it represents the power of music. Formula is as follow:

$$Intesity_n = 20\lg\sqrt{\frac{\sum_{i=1}^{N} S_n^2(i)}{N}} \tag{1.1}$$

- Brightness

Brightness is the centre of mass for spectrum, indicating the spectrum shape and music frequency.

$$Brightness_n = \frac{\sum_{j=1}^{M} F_n(j) W_n{}^2(j)}{\sum_{j=1}^{M} W_n{}^2(j)} \tag{1.2}$$

- Bandwidth

Bandwidth is related to Brightness. Brightness shows the centre of mass for spectrum, while Bandwidth indicates the coverage of spectrum surrounding the centre of mass.

$$Bandwidth_n = \sqrt{\frac{\sum_{j=1}^{M} (F_n(j) - Brightness_n)^2 W_n{}^2(j)}{\sum_{j=1}^{M} W_n{}^2(j)}} \tag{1.3}$$

- Roll off:

Roll off is also an index of spectrum shape. Accumulate power of frequency from low to high until 95% of the whole spectrum power. Frequency at this point is called roll off.

$$\sum_{i=1}^{Rolloff_n} W_n{}^2(i) = 0.95 \sum_{l=1}^{M} W_n{}^2(l) \tag{1.4}$$

- Secondary subband energy distribution

Divide subband frequency into 7 equal width intervals named secondary subband. Energy distribution of secondary subbands is an important quality feature of music. Each instrument has its own specific spectrum energy distribution. The experiment takes the first subband [0，625) as example, dividing it into 7 secondary subbands. The following are Frequency areas:

[0,89),[89,178),[178,267),[267,356),[356,445),[448,534), [534,625)

$$TotalEng_n = \sum_{j=1}^{M} W_n{}^2(j) \tag{1.5}$$

$$SubbandEng_n = \sum_{j=K}^{K+L} W_n^{\ 2}(j) \qquad (1.6)$$

$$Subband\_Energy = \frac{SubbandEng_m}{TotalEng_n} \qquad (1.7)$$

Table1.1 Feature values with four window functions

	Int.	Bri.	Ban.	Rol.	Sub.1	Sub.2	Sub.3	Sub.7
Hamming	-24.0309	272.6406	203.9517	667	0.0897	0.0908	0.1038	0.1629
Hanning	-24.3335	274.3158	204.8736	667	0.0899	0.0910	0.1041	0.1660
Blackman	-25.7373	280.3864	207.4845	668	0.0899	0.0910	0.1064	0.1760
Kaiser	-25.4068	279.0003	203.9517	668	0.0899	0.0910	0.1058	0.1738

1.2.2.2 SVM based on Euclidean core

Assume that there are two sets, $A = \{a_1, ..., a_p\}, B = \{b_1, ..., b_q\}$, the Euclidean distance of the two samples is defined as :

$$d(A, B) = \sqrt{\sum_{k=1}^{p} | A_k - B_k |^2} \qquad (1.8)$$

SVM based on Euclidean distance is defined as:

$$f(x, a) = \sum_{i=1}^{N} a_i y_i K(x, x_i) + a_0 \qquad (1.9)$$

Here, $K(x, x_i)$ is the core function, $K(x, x_i) = \exp(-\alpha d(x, x_i))$. $d(x, x_i)$ is Euclidean distance between sample X and sample X_i . $a = (a_1, a_2, ..., a_N)$ is parameter vector gain from training samples.

We select 150 bad samples marked as -1 and 150 good samples marked as 1. These 300 samples compose SVM training group. Input test data which has been processed with four window functions and the classification results could be generated. In experimental result, 1 is defined as good music quality, 0 is defined as normal music quality and -1 is defined as poor music quality.

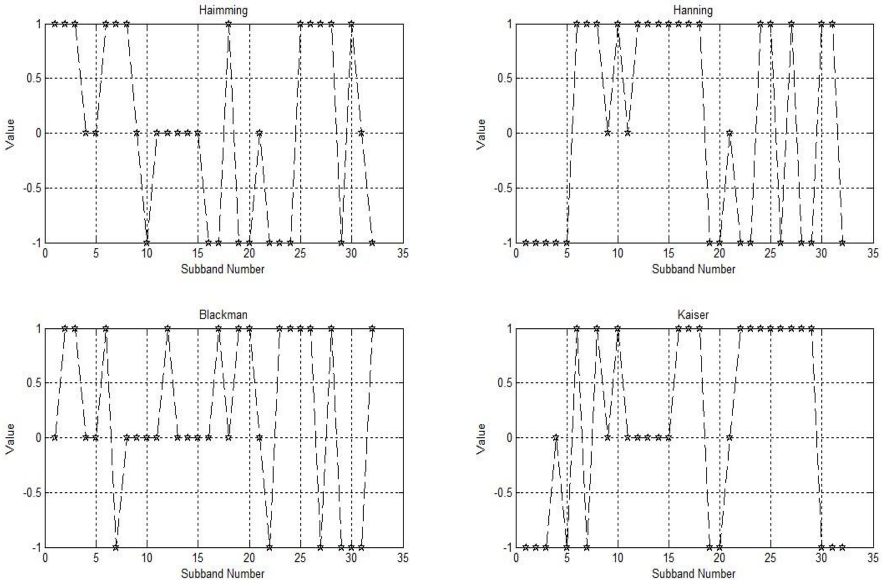

Fig.1.4 Classification Result

The experimental result demonstrates Hamming window contributes to excellent hearing effect in low frequency bands; Hanning window has good performance in high frequency bands; Blackman window is suitable in all bands. Kaiser window helps in smooth transition between adjacent bands in high frequency bands.

The result further improves our test for index M and K, providing powerful support for window selection.

1.3 Conclusion：

This paper discussed the window function choice in Phase Vocoder algorithm. Based on mp3 encoding principle, Human auditory spectrum is divided into 32 equal width subbands. Experiments are operated for each subband with four well known window functions: Hamming window, Hanning window, Blackman window and Kaiser window. Quantitative analysis contains two major steps: First, combined with acoustic sensory pleasantness model, the experiment puts convergence as the index to draw preliminary conclusion on window selection in low, middle and high frequency area. Second, experiment defined and abstracted features from the testing music. And based on SVM, we use 300 samples as training input to form determining function for window evaluation. During the experiment, normalization is used several times to compress data, develop processing speed. Experiment shows, for symphony phase vocoder, Hamming window contributes to precision and convergence in low frequency bands; Hanning window has good performance in middle and high frequency bands; Blackman window get average result in most subbands; Kaiser window proposes the connection between bands in stacking process and being a good choice in high frequency bands.

1.4 References

1. J. L. Flanagan and R. M. Golden, “Phase vocoder,” Bell Syst. Tech. J., vol. 45, pp. 1493–1509, Nov. 1966; also in Speech Analysis, R. W. Schaefer and J. D. Markel, Eds. New York: IEEE Press, 1979.

2. S.S.Abeysekera. K.P.Padhi, J.Absar and S.George. “Investigation of different frequency estimation techniques using the phase vocoder”. Proceedings - IEEE International Symposium on Circuits and Systems

3. Fricke, J.Robert; Cook, George E. “Real-time windowing in imaging radar using FPGA technique” IEEE, New York, NY

4. Ponomaryov, Volodymyr I. Escamilla-Hernandez, Enrique. “Real-time windowing in imaging radar using FPGA technique”. Proceedings of SPIE - The International Society for Optical Engineering

4.Olson, Harry F. (1967). Music, Physics and Engineering. Dover Publications. p. 249. ISBN 0486217698.

6. Ruzanski, Evan P. “Effects of MP3 encoding on the sounds of music” Institute of Electrical and Electronics Engineers Inc.

7. Fastl,H. (Technical Univ Munich, Munich, Germany) “Psychoacoustics of sound-quality evaluation” Acta Acustica Stuttgart

8. J.L. Flanagan and R.M. Golden, “Phase vocoder,” Bell Syst. Tech. J., vol. 45, pp. 1493–1509, Nov 1966.

9.Miller S. Puckette and Judith C. Brown. “Accuracy of Frequency Estimates Using the Phase Vocoder” IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 6, NO. 2, MARCH 1998

10. ISO/IEC JTC1/SC29/WG11 MPEG, IS11172-3 “Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5Mbits/s, Part 3: Audio” 1992

11. J. Laroche and M. Dolson, “Improved Phase Vocoder Time Scale Modification of Audio”, IEEE Trnrisactioris on Speech arid Audio Processirig, M.iy 1999, vol. 7. no. 3. pg. 323.

12. Y ôiti Suzuki (Research Institute of Electrical Communication, Tohoku University: Japan) Precise and Full-range Determination of Two-dimensional Equal Loudness Contours

13. 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