# The Quality Comparison of Audio Time Shift Effects Using Different Sub-Bands Window Function and RBF Classify

**Fan Lin, Jianbin Xiahou, Hao Wang, Yue Wamg**

*School of Software, Xiamen University, Xiamen, China*
*Corresponding author : jbxiahou@xmu.edu.cn*

Corresponding author is Jianbin Xiahou

Abstract
The phase vocoder has long been a well-established tool in audio 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. 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 used in digital audio singal processing technology. 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 into 32 equal width sub-bands. Band Bass Filter is used to implement audio input division. Then, for each sub-band, four window functions are separately used to intercept sound. This paper takes acoustic sensory pleasantness model for sound evaluations. Two indicators of frequency accuracy are discussed. Moreover, the RBF neural network method is introduced to classify suitable and unsuitable window for each sub-band. Finally, we conclude the suitable frequency scale for each of the four window functions.
Keywords: PHASE VOCODER; WINDOWING; MP3; RBF NEURAL NETWORK

## 1. Introduction

The real-time pitch shifting [1] 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 [2][3] 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 [4] frequency is between 20 Hz and 20 kHz, the sine waves frequencies are within this range. Changing the audio pitch is to change the wave frequency which consist 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) [5] is widely used to intercept audio into small pieces of signal [6]. 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 [7].

## 2. Experiment

### 2.1. Sub-Band Filtering

With the help of mp3 encoding principle, human hearing frequency range is separated into 32 equivalent width sub-bands. Each sub-band will be operated for windowing analysis. The experiment uses multiple band-pass filters for audio division shown in Fig. 1.

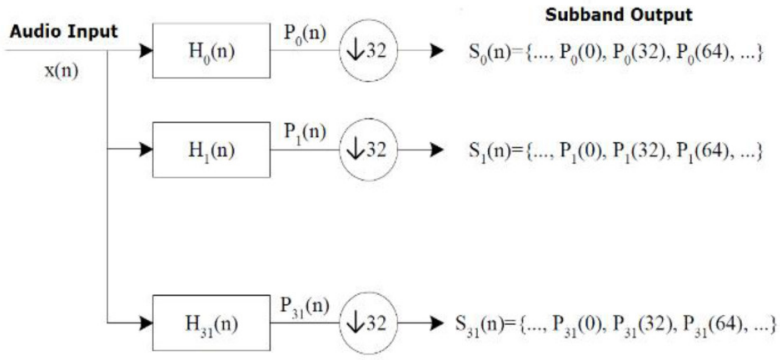


**Figure 1.** Audio input is divided into 32 sub-band output through multiple band-pass filter

Multiple band-pass filter mathematic formula:

$$H_i[n] = h[n] \times \cos\left[\frac{\pi \times (2 \times i + 1) \times (n - 16)}{16}\right] \quad where\ i = 0\sim31,\ n = 0\sim511 \tag{1}$$

$h_i[n]$ is a Low-pass filter.

Taking the signal with 40 kHz sampling as the sample, Each sub-band is 688Hz in width. The center frequency of sub-band $i$ is $20\text{k} \times (2 \times \text{i} + 1) \div (2 \times 32)$.

### 2.2. Acoustic sensory pleasantness

Human acoustic sensation is influenced by factors such as roughness, sharpness, pitch and power of the sound [8]. According to the feature of human hearing system, the audible response to different frequency is remarkably different.

There is a lowest point in equal loudness contour [9]. Its frequency range is about 3k-7k Hz, which is the most sensitive part for human hearing. When the audio frequency is higher or lower than this range, hearing sensitivity is gradually decreasing. But, other frequency bands will still influence hearing sensitivity.

10k-16k Hz is the highest frequency band in symphony, which is also full of expression. However, if this band is over strong, the music will produce burr-like jarring high frequency noise. So we should make certain reproduction on this band.

7k-10k Hz plays an important role on sound clarity and transparency. If the music is lack of this frequency band, the quality will be very dull. If the music consists of too much component on this band, the dental voice will become too sharp and too obvious, which makes people feel tire of hearing.

Too much power on frequency band less than 3k Hz causes the audio being monotonous and plain.

Therefore, we should stress on sensitive frequency band while processing audio.

### 2.3. Comparison of frequency accuracy

The processes of phase synthesis includes: Fourier transform with windowing signal, frequency and phase adaptation, comprehensive windowing and output stacking [10]. This paper uses Hamming window, Hanning Window Function, Blackman[11][12] Window Function, Kaiser Window Function as the operator in discrete Fourier transform and compare the effect of frequency leakage suppression of these window functions in different audio frequency samples.

Blackman Window

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

Hamming Window

$$w_m(n) = \begin{cases} 0.54 - 0.46\cos\frac{2\pi n}{M} & 0 \le n < M \\ 0 & \text{others} \end{cases} \tag{3}$$

Hanning Window

$$w_m(n) = \begin{cases} 0.5\left[1 - \cos\left(\frac{2\pi n}{M+1}\right)\right] \\ 0 \end{cases} \tag{4}$$

Kaiser Window

$$w_m(n)=\begin{cases}\dfrac{I_0\left[\beta\sqrt{1[-\left[\left(2n/(M-1)-1\right)^2\right]}\right]}{I_0[\beta]}\\ 0\end{cases} \quad (5)$$

**2.4. Index of energy concentration**

After spectrum normalization, the experiment uses the ratio related to integration between cut-off frequencies and integration of all spectrum frequency as the index of energy concentration. Function window with more concentrated main lobe is more suitable as Convolution window in Fourier transformation. We set 100Hz besides the main frequency as the cut-off frequency and define spectrum interval within cut-off frequencies as main lobe.

$$\mathrm{M}=\frac{\int_{f-100}^{f+100}xdx}{\int_{0}^{20000}xdx} \quad (6)$$

Here, M is the ratio shown in Tab.1. f is testing frequency. $\int_0^{20000}xdx$ stands for the integration of all spectrum frequency. $\int_{f-100}^{f+100}xdx$ stands for integration between cut-off frequencies.

In order to compare the convergence speed of the side-lobe, we take the slope from the top point of the testing frequency to the cut-off point as the index of convergence. Larger slope indicates the more concentrated main-lobe energy and less side-lobe. That is to say the window function has better performance on restraining frequency leakage.

$$K=\frac{y-y_0}{x-x_0} \quad (7)$$

Here, $K$ is the slope shown in Tab.2. $(x,y)$ stands for the top point of testing frequency, and $(x_0,y_0)$ stands for the cross point of the cut-off frequency and audio strength wave. $y=1$ after normalization.

In this experiment, after windowing and Fourier transformation, we normalize the resulting data, setting the max audio strength as 1 and making the strength value between 0 and 1.

**2.5. Windowing graphic analysis**

According to the acoustic sensory pleasantness model above, we choose three representative sub-bands separately in low, middle and high frequency for analysis.

The experiment takes the first sub-band whose central frequency is 312.5Hz as the sample of low frequency sub-bands, as shown in Fig.2. This sub-band 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.

**Tab. 1.** M for each sub-band and each window function

| Subband | Center Frequency | Hamming | Hanning | Blackman | Kaiser |
|---|---|---|---|---|---|
| 1 | 312.5 | 57.1150 | 56.9598 | 56.1026 | 55.8947 |
| 2 | 937.5 | 57.2669 | 57.0952 | 56.2407 | 56.0380 |
| 3 | 1562.5 | 57.2639 | 57.1019 | 56.2475 | 56.0423 |
| 4 | 2187.5 | 57.2786 | 57.1099 | 56.2557 | 56.0526 |
| 5 | 2812.5 | 57.2725 | 57.1084 | 56.2541 | 56.0492 |
| 6 | 3437.5 | 57.2803 | 57.1125 | 56.2583 | 56.0551 |
| 7 | 4062.5 | 57.2753 | 57.1104 | 56.2562 | 56.0514 |
| 8 | 4687.5 | 57.2809 | 57.1135 | 56.2594 | 56.0560 |
| 9 | 5312.5 | 57.2767 | 57.1115 | 56.2573 | 56.0525 |
| 10 | 5937.5 | 57.2811 | 57.1141 | 56.2599 | 56.0566 |
| 11 | 6562.5 | 57.2777 | 57.1121 | 56.2580 | 56.0533 |
| 12 | 7187.5 | 57.2813 | 57.1145 | 56.2603 | 56.0569 |
| 13 | 7812.5 | 57.2784 | 57.1126 | 56.2584 | 56.0538 |
| 14 | 8437.5 | 57.2814 | 57.1147 | 56.2606 | 56.0572 |
| 15 | 9062.5 | 57.2790 | 57.1130 | 56.2588 | 56.0542 |
| 16 | 9687.5 | 57.2814 | 57.1150 | 56.2608 | 56.0574 |
| 17 | 10312.5 | 57.2795 | 57.1134 | 56.2592 | 56.0546 |
| 18 | 10937.5 | 57.2814 | 57.1152 | 56.2610 | 56.0576 |
| 19 | 11562.5 | 57.2800 | 57.1137 | 56.2595 | 56.0550 |
| 20 | 12187.5 | 57.2814 | 57.1154 | 56.2612 | 56.0577 |
| 21 | 12812.5 | 57.2805 | 57.1140 | 56.2599 | 56.0554 |
| 22 | 13437.5 | 57.2813 | 57.1155 | 56.2614 | 56.0579 |
| 23 | 14062.5 | 57.2812 | 57.1144 | 56.2603 | 56.0558 |
| 24 | 14687.5 | 57.2812 | 57.1157 | 56.2616 | 56.0580 |
| 25 | 15312.5 | 57.2819 | 57.1148 | 56.2607 | 56.0563 |
| 26 | 15937.5 | 57.2809 | 57.1159 | 56.2618 | 56.0582 |
| 27 | 16562.5 | 57.2830 | 57.1154 | 56.2613 | 56.0569 |
| 28 | 17187.5 | 57.2801 | 57.1160 | 56.2619 | 56.0581 |
| 29 | 17812.5 | 57.2845 | 57.1159 | 56.2618 | 56.0576 |
| 30 | 18437.5 | 57.2765 | 57.1148 | 56.2606 | 56.0566 |
| 31 | 19062.5 | 57.2841 | 57.1123 | 56.2581 | 56.0544 |
| 32 | 19687.5 | 57.1755 | 57.0198 | 56.1638 | 55.9573 |

The experiment takes the 12th sub-band whose central frequency is 7187.5 Hz as the sample of middle frequency sub-bands, as shown in Fig.3. Based on the acoustic sensory pleasantness model, human hearing is sensitive 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 Function has the highest slope, namely the fastest convergence

**Tab. 2.** K for each subband and each window function

| Subband | Center Frequency | Hamming | Hanning | Blackman | Kaiser |
|---|---|---|---|---|---|
| 1 | 312.5 | 13.5464 | 13.6968 | 15.4662 | 10.9290 |
| 2 | 937.5 | 29.3509 | 30.5844 | 28.3241 | 25.8003 |
| 3 | 1562.5 | 13.5374 | 13.6872 | 15.4567 | 10.9192 |
| 4 | 2187.5 | 29.3502 | 30.5832 | 28.3224 | 25.7994 |
| 5 | 2812.5 | 13.5373 | 13.6870 | 15.4565 | 10.9189 |
| 6 | 3437.5 | 29.3502 | 30.5830 | 28.3220 | 25.7992 |
| 7 | 4062.5 | 13.5372 | 13.6869 | 15.4565 | 10.9188 |
| 8 | 4687.5 | 29.3501 | 30.5829 | 28.3218 | 25.7992 |
| 9 | 5312.5 | 13.5372 | 13.6869 | 15.4564 | 10.9188 |
| 10 | 5937.5 | 29.3501 | 30.5829 | 28.3217 | 25.7992 |
| 11 | 6562.5 | 13.5372 | 13.6869 | 15.4564 | 10.9188 |
| 12 | 7187.5 | 29.3501 | 30.5828 | 28.3217 | 25.7991 |
| 13 | 7812.5 | 13.5372 | 13.6869 | 15.4564 | 10.9188 |
| 14 | 8437.5 | 29.3501 | 30.5828 | 28.3216 | 25.7991 |
| 15 | 9062.5 | 13.5372 | 13.6869 | 15.4564 | 10.9188 |
| 16 | 9687.5 | 29.3501 | 30.5828 | 28.3216 | 25.7991 |
| 17 | 10312.5 | 13.5372 | 13.6869 | 15.4564 | 10.9188 |
| 18 | 10937.5 | 29.3501 | 30.5828 | 28.3215 | 25.7991 |
| 19 | 11562.5 | 13.5372 | 13.6869 | 15.4564 | 10.9187 |
| 20 | 12187.5 | 29.3502 | 30.5827 | 28.3215 | 25.7991 |
| 21 | 12812.5 | 13.5372 | 13.6869 | 15.4564 | 10.9187 |
| 22 | 13437.5 | 29.3502 | 30.5827 | 28.3214 | 25.7991 |
| 23 | 14062.5 | 13.5372 | 13.6869 | 15.4564 | 10.9187 |
| 24 | 14687.5 | 29.3502 | 30.5827 | 28.3213 | 25.7991 |
| 25 | 15312.5 | 13.5372 | 13.6869 | 15.4564 | 10.9187 |
| 26 | 15937.5 | 29.3502 | 30.5827 | 28.3213 | 25.7991 |
| 27 | 16562.5 | 13.5373 | 13.6869 | 15.4565 | 10.9188 |
| 28 | 17187.5 | 29.3503 | 30.5827 | 28.3211 | 25.7991 |
| 29 | 17812.5 | 13.5374 | 13.6870 | 15.4566 | 10.9188 |
| 30 | 18437.5 | 29.3505 | 30.5827 | 28.3209 | 25.7993 |
| 31 | 19062.5 | 13.5382 | 13.6876 | 15.4572 | 10.9194 |
| 32 | 19687.5 | 29.3573 | 30.5878 | 28.3236 | 25.8053 |

speed and a comparatively concentrated M value. Besides, Blackman window also has a good performance in this sub-band. But Kaiser Window spreads wave power to adjacent area obviously, indicating a bad behavior.

The experiment takes the 24th sub-band whose central frequency is 14687.5Hz as the sample of middle frequency sub-bands, as shown in Fig.4. 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 sub-bands in splicing process. Therefore Hamming window is not recommended for Fourier transform in high frequency sub-bands. On the other hand, other three windows could improve the transition process smoothly and satisfies the requirement for the high frequency sub-bands.

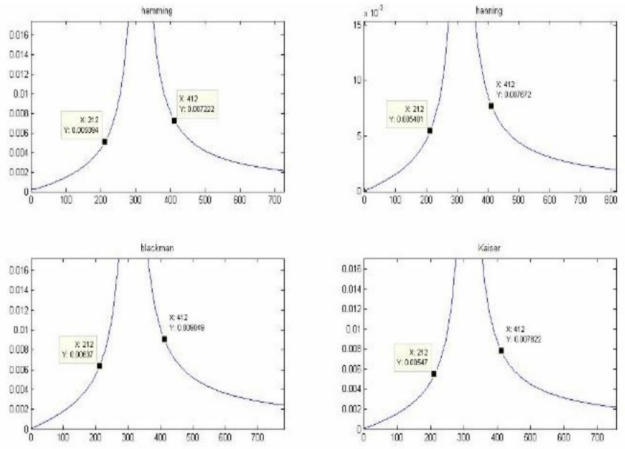


**Figure 2.** Low frequency sub-band windowing analysis

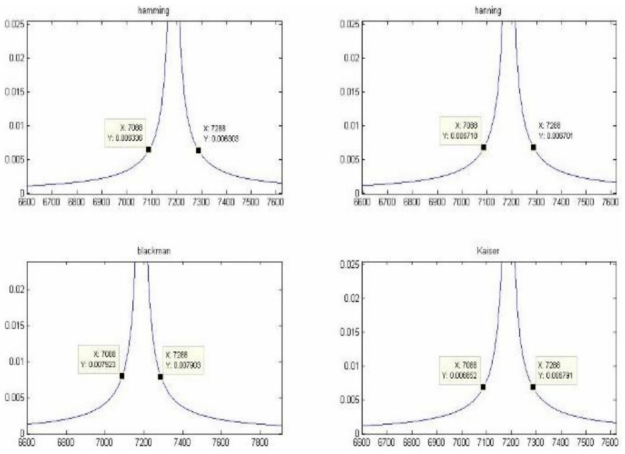


**Figure 3.** Middle frequency sub-band windowing analysis

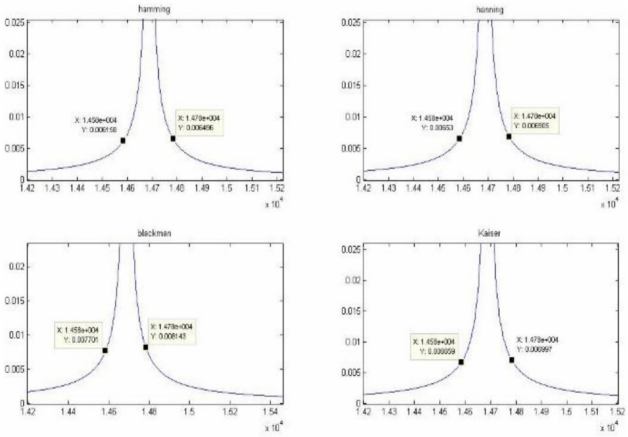


**Figure 4.** High frequency sub-band windowing analysis

### 2.6. Classification of window function based on RBF neural network

Radical Basis Function (RBF) neural network is a learning method that analyzes data and recognizes patterns. Here we use RBF neural network as a method to classify suitable or unsuitable window functions for 32 frequency sub-bands. Given a set of training examples, each marked as suitable or unsuit-

able, the RBF neural network training algorithm will build a model. With the help of this model, any new input can be classified correctly. RBF neural network contributes to the transformation from dot product in multidimensional feature space to core function arithmetic in low dimensional space, voiding multidimensional computing. We use radial basis function based on Euclidean distance.

We define 8 features for music and abstract these features for all training examples and test input before applying RBF neural network.

**2.7. Music feature abstraction:**

In order to evaluate the output music after pitch shifting with quantizing analysis method, we define a series 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:

$$Intesity_n = 20\lg\sqrt{\frac{\sum_{i=1}^{N} S_n^2(i)}{N}} \tag{8}$$

Brightness:

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

$$Brightness_n = \frac{\sum_{j=1}^{M} F_n(j) W_n{}^2(j)}{\sum_{j=1}^{M} W_n{}^2(j)} \tag{9}$$

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.

$$Bandwidth_n = \sqrt{\frac{\sum_{j=1}^{M} (F_n(j) - Brightness_n)^2 W_n{}^2(j)}{\sum_{j=1}^{M} W_n{}^2(j)}} \tag{10}$$

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.

$$\sum_{i=1}^{Rolloff_n} W_n{}^2(i) = 0.95 \sum_{l=1}^{M} W_n{}^2(l) \tag{11}$$

Secondary sub-band energy distribution:

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

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

$$TotalEng_n = \sum_{j=1}^{M} W_n{}^2(j) \tag{12}$$

$$SubbandEng_n = \sum_{j=K}^{K+L} W_n{}^2(j) \tag{13}$$

$$Subband\_Energy = \frac{SubbandEng_m}{TotalEng_n} \tag{14}$$

**Table 3.** Feature values with four window functions

| | Hamming | Hanning | Blackman | Kaiser |
|---|---|---|---|---|
| Int. | -24.0309 | -24.3335 | -25.7373 | -25.4068 |
| Bri. | 272.6406 | 274.3158 | 280.3864 | 279.0003 |
| Ban. | 203.9517 | 204.8736 | 207.4845 | 203.9517 |
| Rol. | 667 | 667 | 668 | 668 |
| Sub.1 | 0.0897 | 0.0899 | 0.0899 | 0.0899 |
| Sub.2 | 0.0908 | 0.091 | 0.091 | 0.091 |
| Sub.3 | 0.1038 | 0.1041 | 0.1064 | 0.1058 |
| Sub.7 | 0.1629 | 0.166 | 0.176 | 0.1738 |

**2.8. RBF neural network based on Euclidean distance**

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 :

$$d(A,B) = \sqrt{\sum_{k=1}^{p} |A_k - B_k|^2} \tag{15}$$

We use Gaussian function:

$$\varphi_i(x) = \exp(-\| x - c_i \|^2 / \theta_i^2), i = 1,2,...,N \tag{16}$$

Here, $x$ is one dimension input vector,. $c_i$ is the center of the $i$ th RBF which has the same dimension with $x$. $\theta_i$ is the width of RBF in the $i$ th neural. $\| x - c_i \|$ is the distance of vector $x - c_i$.

The output of the network is:

$$\hat{y} = f(x) = w_0 + \sum_{i=1}^{N} w_i \exp(-\| x - c_i \|^2 / \sigma^2_i) \qquad (17)$$

We select 150 bad samples marked as -1 and 150 good samples marked as 1. These 300 samples compose RBF neural network 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.

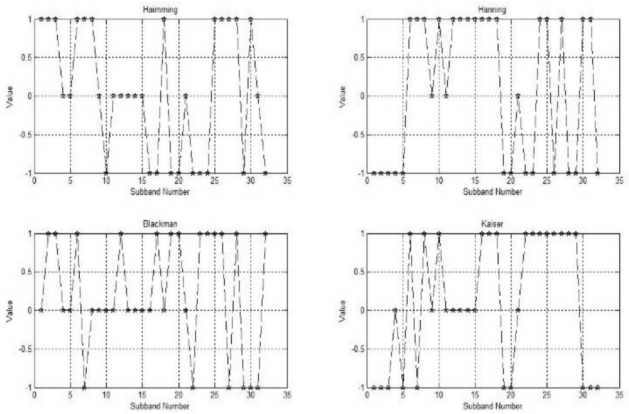


**Figure 5.** Classification Result

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.

**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 sub-bands. Experiments are operated for each sub-band 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 RBF neural network, we use 300 samples as training input to form network 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 sub-bands; Kaiser window proposes the connection between bands in stacking process and being a good choice in high frequency bands.

**References**

1. 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
2. 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.
3. Fastl,H. (Technical Univ Munich, Munich, Germany) “Psychoacoustics of sound-quality evaluation” Acta Acustica Stuttgart
4. Olson, Harry F. (1967). Music, Physics and Engineering. Dover Publications. p. 249. ISBN 0486217698.
5. J.L. Flanagan and R.M. Golden, “Phase vocoder,” Bell Syst. Tech. J., vol. 45, pp. 1493–1509, Nov 1966.
6. Ruzanski, Evan P. “Effects of MP3 encoding on the sounds of music” Institute of Electrical and Electronics Engineers Inc.
7. 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
8. 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.
9. Yôiti Suzuki (Research Institute of Electrical Communication, Tohoku University: Japan) Precise and Full-range Determination of Two-dimensional Equal Loudness Contours
10. 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
11. Fricke, J.Robert; Cook, George E. “Real-time windowing in imaging radar using FPGA technique” IEEE, New York, NY

12. Ponomaryov, Volodymyr I. Escamilla-Hernandez, Enrique. “Real-time windowing in imaging radar using FPGA technique”. Proceedings of SPIE - The International Society for Optical Engineering
13. A. R¨obel, “A Shape-Invariant Phase Vocoder for Speech Transformation,” in Proc. Digital Audio Effects (DAFx-10)(2010).
14. C. Sch¨orkhuber and A. Klapuri, “Constant-Q Transform Toolbox for Music Processing,” in Proc. Sound and Music Computing Conference (SMC) (2010).
15. G. A. Velasco, N. Holighaus, M. D¨orfler, and T. Grill, “Constructing an Invertible Constant-Q Transform with Nonstationary Gabor Frames,” in Proc. Digital Audio Effects (DAFx-11) (2011).
16. P. Balazs, M. D¨orfler, F. Jaillet, N. Holighaus, and G. Velasco, “Theory, Implementation and Applications of Nonstationary Gabor Frames,” J. Computational and Applied Mathematics, pp. 236:1481–1496 (2011).
17. I. W. Selesnick, “Wavelet Transform with Tunable Q-Factor,” IEEE Transactions on Signal Processing, vol. 59, no. 8, pp. 3560 (2011).
18. C. Sch¨orkhuber, A. Klapuri, and A. Sontacchi, “Pitch Shifting of Audio Signals Using the Constant-Q Transform,” in Proc. Digital Audio Effects (DAFx-12) (2012).
19. C. Sch¨orkhuber, “Pitch Shifting Using the CQT: Audio Examples,” Available at http:// www.iem.at/ schoerkhuber/ cqt/, accessed July 09, 2012.
20. S. Sofianos, A. Ariyaeeinia, R. Polfreman, and R. Sotudeh, “H-Semantics: A Hybrid Approach to Singing Voice Separation,” J. Audio Eng. Soc, vol. 60, pp. 831–841(2012 Oct.).
21. D. Gunawan and D. Sen, “Separation of Harmonic
Musical Instrument Notes Using Spectro-Temporal Modeling
of Harmonic Magnitudes and Spectrogram Inversion with Phase Optimization,” J. Audio Eng. Soc, vol. 60,pp. 1004–1014 (2012 Dec.).
22. D. Fitzgerald, “Harmonic/Percussive Separation Using Median Filtering,” in Proc. Digital Audio Effects (DAFx-10)(2010).
23. Celia Shahnaz,Wei-Ping Zhu,M. Omair Ahmad.Pitch Estimation Based on a Harmonic Sinusoidal Autocorrelation Model and a Time-Domain Matching Scheme. IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING . 2012
24. S.Gonzalez,M.Brooks."A pitch estimation filter robust to high levels of noise (PEFAC) ". 19th European Signal Processing Conference (EUSIPCO 2011).
25. Liu Yu,Ma Yan. Application of Web and Multimodality to Linguistics Teaching [A]. Proceedings of 2nd International Conference on Science and Social Research (ICSSR 2013) [C]. 2013.
26. Wlodarczyk,Michal and Sekalski, Przemyslaw, Evaluation of Time-Scale Modification Methods for Audio Signals on Mobile Devices with Android OS, 21st International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES), Lublin, POLAND, 2014 PROCEEDINGS OF THE 21ST INTERNATIONAL CONFERENCE ON MIXED DESIGN OF INTEGRATED CIRCUITS & SYSTEMS (MIXDES), p. 451-454, 2014.
27. Driedger, Jonathan; Mueller, Meinard; Ewert, Sebastian, Improving Time-Scale Modification of Music Signals Using Harmonic-Percussive Separation, IEEE SIGNAL PROCESSING LETTERS, p. 105-109, 2014.
28. Bin,Xiao;Yi,Jiang,The comparison of window functions for different subbands in Phase Vocoder, 3rd International Conference on Multimedia Technology (ICMT), Guangzhou, PEOPLES R CHINA, NOV 29-DEC 01, p. 1465-1472, 2013.
29. Sondergaard, Peter L, Efficient Algorithms for the Discrete Gabor Transform with a Long Fir Window, JOURNAL OF FOURIER ANALYSIS AND APPLICATIONS, p. 456-470, 2012.
30. Bayram, Ilker; Selesnick, Ivan W, Bayram, Ilker; Selesnick, Ivan W, EEE TRANSACTIONS ON SIGNAL PROCESSING, p. 6251-6256, 2011.