

Research Article

Feature Extraction and Analysis Method of Trombone Timbre Based on CNN Model

Yanjun Wang 

School of Music, Shandong College of Arts, Jinan 250014, China

Correspondence should be addressed to Yanjun Wang; z00922@sdca.edu.cn

Received 10 August 2022; Revised 30 August 2022; Accepted 13 September 2022; Published 23 September 2022

Academic Editor: Shahid Hussain

Copyright © 2022 Yanjun Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the accuracy of trombone timbre feature extraction, this paper combines the CNN model to construct a trombone timbre feature extraction model and summarizes the principle of trombone timbre signal. Moreover, this paper deduces the parameters of the trombone timbre signal and the corresponding network model and uses mathematical expressions to model the trombone timbre signal, which is convenient for theoretical analysis and processing of the trombone timbre signal. In addition, this paper provides a detailed discussion of time-frequency analysis techniques, including their advantages and limitations, which provide an algorithmic basis for working with trombone timbre signals. It can be seen that time-frequency analysis technology still has great advantages in trombone timbre signal processing. Finally, the simulation results show that the trombone timbre feature extraction method based on the CNN model proposed in this paper can effectively identify the trombone timbre in various musical performances.

1. Introduction

At the beginning of its invention, the trombone was divided into treble trombone, alto trombone, sub trombone, and bass trombone. However, with the continuous invention and improvement of musical instruments, the treble trombone and the alto trombone in the trombone type were gradually replaced by other instruments, and only the infratone trombone and the bass trombone remained. With the continuous development of the times, the trombone has evolved into a relatively mature musical instrument and is widely loved by many music lovers because of its unique timbre charm. Moreover, it plays an important and irreplaceable role in the performance of wind instruments [1].

Timbre refers to the feeling that sound brings to people, and it can also be understood as the quality of sound. Timbre has the function of judging the performance of musical instruments, such as the effective judgment of many factors including the smoothness, strength, and weakness, and emotion of the sound produced when the relevant musical instrument is played. The understanding of timbre can be

achieved by observing the feeling of color with the naked eye, and the ear will also form a corresponding impression when hearing the sound, that is, “timbre.” It should be noted that the realization of beautiful timbre cannot be judged only from a certain angle, such as the thickness and brightness of the sound, but needs to be comprehensively interpreted in combination with different scenes [2]. This is like in the painting process: the perfect painting cannot be achieved only through a single color and line, but through different colors to show the layering and artistic charm of the painting. In the trombone performance, it is necessary to use different timbres according to the different types of pieces to be played, instead of using a single timbre for performance. It is recognized as a deduction that achieves beautiful timbre [3]. For example, performing classical repertoire needs to be accompanied by a calm and vigorous tone; performing romantic music requires a bright and cheerful tone. At present, modern musical instruments also use different timbre interpretation methods in the performance process, and it is still necessary to combine the style of music for correct interpretation. Therefore, to realize the beautiful

timbre of the trombone, it must be feasible under the premise of meeting the requirements of music and understanding the concept [4].

In the performance of the trombone, the performance level and ability of the individual will have the most direct impact on the performance. Therefore, the key to whether the trombone can play a beautiful timbre depends on the performer. Although the ability to play the trombone can be acquired through a long period of hard training, it needs to be clear that the correct playing method must be mastered in order to achieve the perfect interpretation of the trombone [5]. Learning the trombone requires the use of the correct mouth shape and breath, and further study on this basis is meaningful. Under normal circumstances, the smile-pressed mouth shape will produce a relatively bright and thin tone when playing the trombone and a lack of control over the intonation and volume, resulting in a high overall tone. Especially when performing classical music or romantic music, it is especially sharp, not round enough, and difficult to coordinate with other musical instruments [6]. The use of a too concentrated mouth shape will produce a relatively dull tone, which is not bright enough and too high-pitched compared to the interpretation of the smile-pressed mouth shape. Because the mouth shape cannot be used flexibly in performance, the performance of gorgeous music or music that requires higher performance skills does not have sufficient performance expression [7]. Therefore, studies have shown that using a mouth shape that cancels each other out with a smile and tuck is by far the correct mouth shape for trombone playing. Because of its flexibility and power, it can fully meet the performance requirements of various pieces of music and can achieve a more ideal effect of trombone performance [8].

In addition to the shape of the mouth, the vibration of the lips and the speed of the airflow caused by breathing also have a major impact on the timbre. In the use of breathing in wind instruments, players often use two methods, one is chest breathing and the other is abdominal breathing. Chest breathing makes the performer easily fatigued due to too little inhalation during the performance, and it is difficult to achieve good breath regulation and strength support in the high-pitched or low-pitched part of the music [9]. Abdominal breathing mainly relies on the breath force of the abdomen to achieve performance. Although the performer can achieve positive movement to generate breath during the performance, it cannot effectively increase the inhalation value and maximize the inhalation. At present, thoracic-abdominal breathing is the most scientific playing method used in wind instrument performance, especially trombones. The muscles of breathing can not only maximize breathing but also control the breathing smoothly, without causing fatigue to the performer [10].

In order to develop the bass trombone to a greater extent in the art of playing, we must learn and master the bass trombone playing techniques. The following will focus on some of the factors that the player can control, such as how to use the mouth skills in shape, breathing method, tongue movement, pronunciation, etc., to make the bass trombone sound more perfect and harmonious [11].

The trombone is different from the keyed instrument. It does not have the flexible control of the keyed instrument, and it is also unrestrained and bold with the percussion instrument, but it can play a sound with its own personality, showing its own unique charm in performance. Moreover, the function of tongue movement is an important expression technique for trombone players. In trombone performance, we can usually hear music with a very fast rhythm. This is to use tongue movement and speed to play. Yes, we should practice a lot of tongue movement skills on the basis of mastering basic skills, such as speeding up the speed of the tongue on the premise of expressing each note neatly and clearly [12]. All in all, we can only improve in tongue movement skills after mastering the method and practicing a lot, and only through hard work can we achieve something in basic skills. Diligently practice the basic skills and learn from the simple to the complex order, and there will be a higher level breakthrough in the performance of the trombone [13].

The lip shape of the bass trombone is very important, it is related to the sound quality and sound quality. The shape of the mouth plays an important and decisive role in the sound quality of the bass trombone. Bass trombone performance relies on the coordination of mouth shape and teeth and other organs. Mouth shape is the most basic learning stage, and bass trombone players must ensure a correct mouth shape [14]. It is necessary to practice patiently and make continuous progress because the innate conditions are different, so the ability to use and control the mouth shape is also decidedly different. We must analyze this situation based on the actual situation [15].

Whether the music is pleasing to the ear or not, the pitch plays an important decision. Pitch specifically refers to the high-pitched sound produced during singing and musical instrument performance that can match the high-pitched sound of a certain rate. Therefore, it is said that pitch determines the basic conditions for playing beautiful, pleasant, and pleasant music. There are many people in life who sing songs that feel particularly beautiful, and the real reason is the pitch [16]. As far as musical instruments are concerned, trombone and pitch have an inseparable and important relationship. Trombone mainly relies on the movement of the telescopic tube to adjust the treble, so it is not easy to correctly control the pitch issue. It takes hard work and time to do the right exercise and inquiry. And if we can usually listen to some excellent performances, it can also help us to improve the pitch [17].

This paper combines the CNN model to construct the trombone timbre feature extraction model to improve the training and learning effect of the trombone and promote the effect of the trombone in playing.

2. Trombone Timbre Frequency Communication Overview

2.1. Trombone Timbre Frequency Communication Principle and Characteristics. As shown in Figure 1, the trombone timbre frequency communication process first generates the original information data from the transmitting end and then performs the first baseband modulation through the

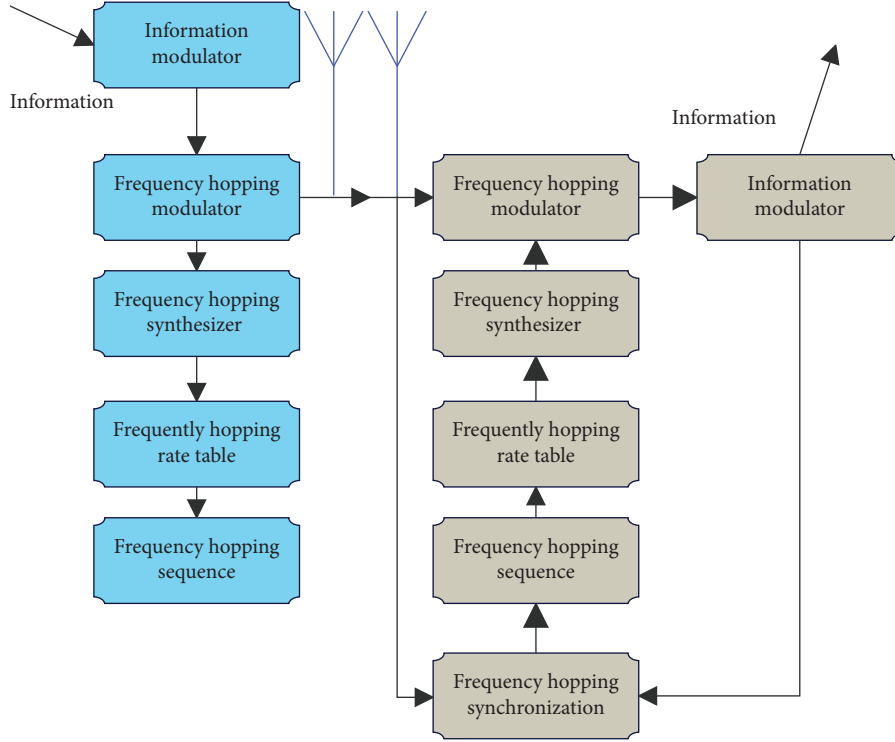


FIGURE 1: Block diagram of the working principle of the trombone timbre frequency communication system.

information modulator. At the same time, the trombone timbre frequency sequence is generated under the control of the pseudorandom sequence, and then the frequency hopping table is synthesized through a specific mapping relationship, so as to control the frequency synthesizer to select the local carrier according to the corresponding rules. After that, the baseband signal is multiplied by the trombone timbre frequency modulator to achieve the purpose of corresponding frequency shifting, that is, the carrier wave. Finally, the frequency band information is radiated into the air through the transmitting antenna.

According to whether the frequency hopping is the same time reference, it is divided into synchronous network and asynchronous network. According to whether the trombone timbre frequency collides at the same time, it is divided into orthogonal network and nonorthogonal network. In general, due to the asynchronous network, it is difficult to avoid the collision of trombone timbre frequencies. Therefore, the trombone timbre frequency networking methods are generally divided into three categories according to the abovementioned situation. A schematic diagram of each network model is shown in Figure 2.

2.2. Mathematical Model of Trombone Timbre Frequency Signal. The trombone timbre frequency signal is a

nonstationary signal whose carrier frequency varies with time. The variation law of its carrier frequency is controlled by pseudorandom sequences, such as m-sequence, gold sequence, and so on.

The regularity of the trombone timbre frequency signal is generally observed through the time-frequency diagram. On a time-frequency diagram, the trombone timbre frequency signal appears as a line that varies in the time-frequency dimension. The time-frequency diagram of a single trombone timbre frequency signal is shown in Figure 3. The horizontal axis of the time-frequency diagram is time, and the vertical axis is frequency. Therefore, from the time-frequency diagram, parameters such as time-hopping, period-hopping, and frequency sets of the trombone timbre frequency signal can be clearly observed. We assume that M trombone timbre frequency signal segments are received at observation time T . There are K complete signal segments, the hopping period is T_h , and the carrier frequency is f_k ($k = 1, 2, \dots, K$). Incomplete signals are mainly the beginning and the end of the two parts. Among them, the start duration is τ_s , the carrier frequency is f_s , the end duration is τ_e , and the carrier frequency is f_e . Then, the expression for the signal throughout the observation time is given as follows:

$$s(t) = a(t) \times \left\{ \text{rect}\left(\frac{t}{\tau_s}\right) \exp(j2\pi f_s) + \sum_{k=2}^{K-1} \text{rect}\left(\frac{t - (k-1)T_h - \tau_s}{T_h}\right) \exp(j2\pi f_k) + \text{rect}\left(\frac{t - (K-1)T_h - \tau_s}{\tau_e}\right) \exp(j2\pi f_e) \right\}. \quad (1)$$

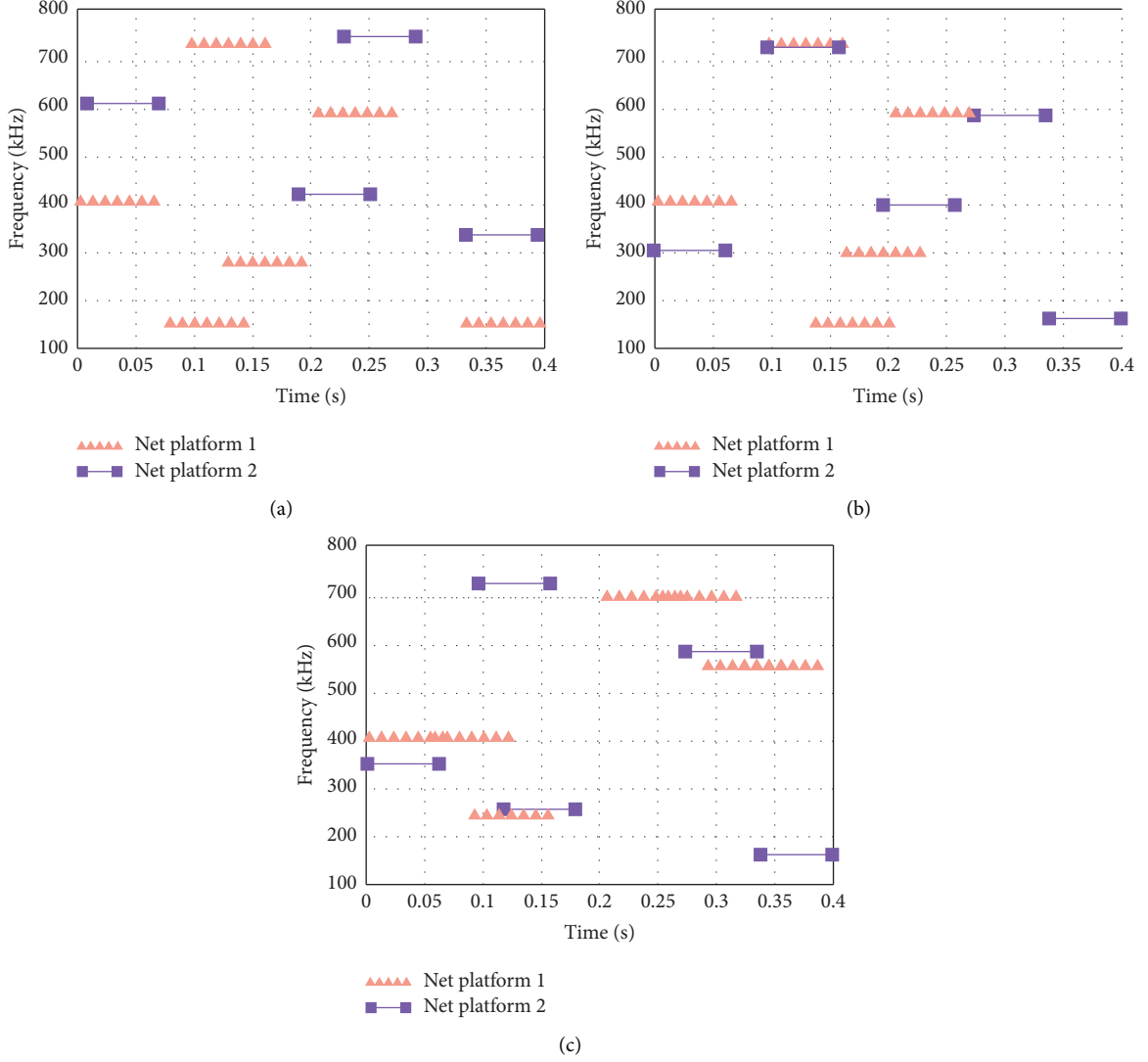


FIGURE 2: Three types of networking. (a) Synchronous orthogonal networking, (b) synchronous nonorthographic networking, and (c) asynchronous nonorthogonal networking.

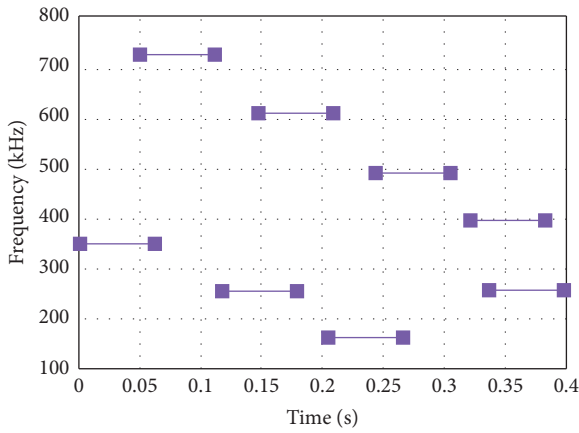


FIGURE 3: Schematic diagram of a single trombone timbre frequency signal.

Among them,

$$\text{rect}\left(\frac{t}{T}\right) = \begin{cases} 1 & 0 < t \leq T, \\ 0 & \text{others,} \end{cases} \quad (2)$$

where $a(t)$ represents the complex envelope of the observed trombone timbre frequency signal.

The signal received in the real communication environment is not only a single trombone timbre frequency signal but is usually mixed with other trombone timbre frequency signals, as well as interference signals, such as fixed frequency signals, sweep frequency signals, and so on. Therefore, in the case of single-antenna reception, it is assumed that N trombone timbre frequency signals are received from the air, and the interference is additive interference, so the expression for single-antenna multi-trombone timbre frequency reception is given as follows:

$$y(t) = \sum_{n=1}^N s_n(t) + n(t). \quad (3)$$

In formula (3), s_n ($n = 1, 2, \dots, N$) is the trombone timbre frequency signal, and $n(t)$ is the sum of various disturbances and noises. Of course, in addition to single-antenna reception, there is also array antenna reception, mainly using linear uniform linear array and uniform circular array. The purpose of applying multiple antennas is to use the signal arrival delay to estimate parameters such as the incident direction angle, which is convenient for the later blind source separation problem. However, this article is mainly a single-antenna system, so it will not be discussed for now.

2.3. Time-Frequency Analysis Technology of Trombone Timbre Frequency Signal. The short-time Fourier transform (STFT) is a classical linear transform proposed by Gabor. It mainly uses $h(t)$ for windowing processing on the basis of Fourier transform. Moreover, each segment of the signal segmented by the window function is considered to be stable, and then the window function is continuously shifted, and the Fourier transform is performed on each segment of the signal, and finally, all the transforms are superimposed. Because the trombone timbre frequency signal $s(t)$ is a nonstationary signal, it has a good effect on STFT processing, and its continuous-time expression is given as follows:

$$\text{STFT}_x(t, f) = \int_{-\infty}^{+\infty} s(\tau) h^*(\tau - t) e^{-j2\pi f\tau} d\tau. \quad (4)$$

The discrete expression is given as follows:

$$\text{STFT}(m, n) = \sum_{k=0}^{N-1} s(m+k) h(k) e^{-j(2\pi/N)kn}, \quad n = 0, 1, \dots, N-1. \quad (5)$$

Among them, m is the number of sampling points in the time dimension, n is the number of sampling points in the frequency dimension, $h(k)$ is the discrete window function, and $\text{STFT}(n, m)$ represents the amplitude at the corresponding coordinate point. It can be seen from the abovementioned formula that STFT also has the following properties:

(1) Time shift property:

$$y(t) = s(t - t_0), \quad (6)$$

$$\text{STFT}_y(t, f) = \text{STFT}_s(t, f - f_0). \quad (7)$$

(2) Frequency shift property:

$$y(t) = s(t) e^{j2\pi f_0 t}, \quad (8)$$

$$\text{STFT}_y(t, f) = \text{STFT}_x(t - t_0, f) e^{j2\pi f_0 t}. \quad (9)$$

For STFT, its time-frequency resolution is mainly affected by the width of the window. If it is a long window, the

resolution in the frequency domain is higher, and the resolution in the time domain is lower; if it is a short window, the result is the opposite. The two contradict each other, which is mainly constrained by the uncertainty principle. Its expression is given as follows:

$$BT \geq \frac{1}{2}. \quad (10)$$

Among them, B is the bandwidth, and T is the time width. Therefore, when analyzing the signal, it is necessary to select the appropriate window function parameters according to different situations. Figure 4 shows the time-frequency diagram of the STFT transformation of the trombone timbre frequency signal processed by different windows, the sampling length of the signal is N , and the type of the window H is the hamming window. The window length of Figure 4(a) is $N/4 + 1$, and the window length of Figure 4(b) is $N/10 + 1$.

It can be seen that when the window length is $N/4 + 1$, the frequency resolution of the trombone timbre frequency signal is high, and when the window length is $N/10 + 1$, the time resolution is high. Therefore, when analyzing the trombone timbre frequency signal, the length of the window depends on the situation. For fast trombone timbre frequency signals, short window processing can be selected. However, for slow trombone timbre frequency signals, long window processing can be selected.

For the case where multiple signals are superimposed, $s_1(t)$ and $s_2(t)$ are added linearly, and the following formula is satisfied:

$$\text{STFT}[as_1(t) + bs_2(t)] = a\text{STFT}[s_1(t)] + b\text{STFT}[s_2(t)]. \quad (11)$$

It can be seen that STFT is a linear transformation. Common ones are Gabor transform and wavelet transform.

Gabor transform is proposed by Gabor in 1946 to represent time-frequency information in the form of a grid in a two-dimensional plane. Compared with the time window of STFT, Gabor is a joint time-frequency window. The Gabor change can be seen as the Gauss window selected by the STFT as the window function.

$$\text{Gabor}(t, f) = \int s(\tau) g_{\text{Gauss}}(\tau - t) e^{j2\pi f\tau} d\tau. \quad (12)$$

When the Gaussian window function is used, the lower bound of the uncertainty theorem can be satisfied. Therefore, the Gabor transform can be regarded as the optimal STFT transform. However, the shape of Gabor transform window function is uniform, unlike STFT, which can choose a variety of window functions to truncate the signal, and the time-frequency window size is fixed. As far as the actual signal processing is concerned, the window length used for signals of different frequencies should be variable. Wavelet change is a linear transformation with different resolutions at different frequencies. Its expression is given as follows:

$$\text{wavelet}(a, b) = \frac{1}{\sqrt{a}} \int f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (13)$$

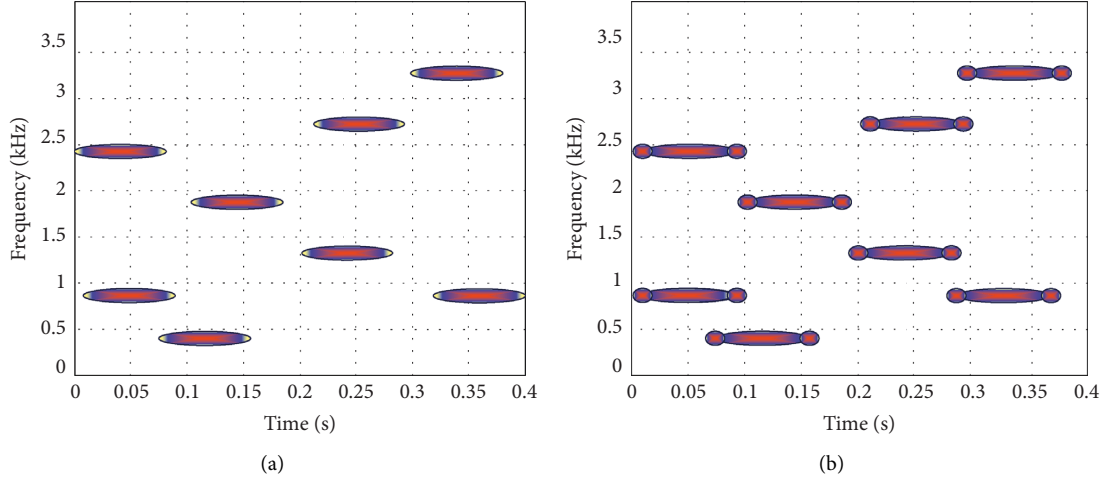


FIGURE 4: STFTs with different window lengths. (a) $H = N/4 + 1$. (b) $H = N/10 + 1$.

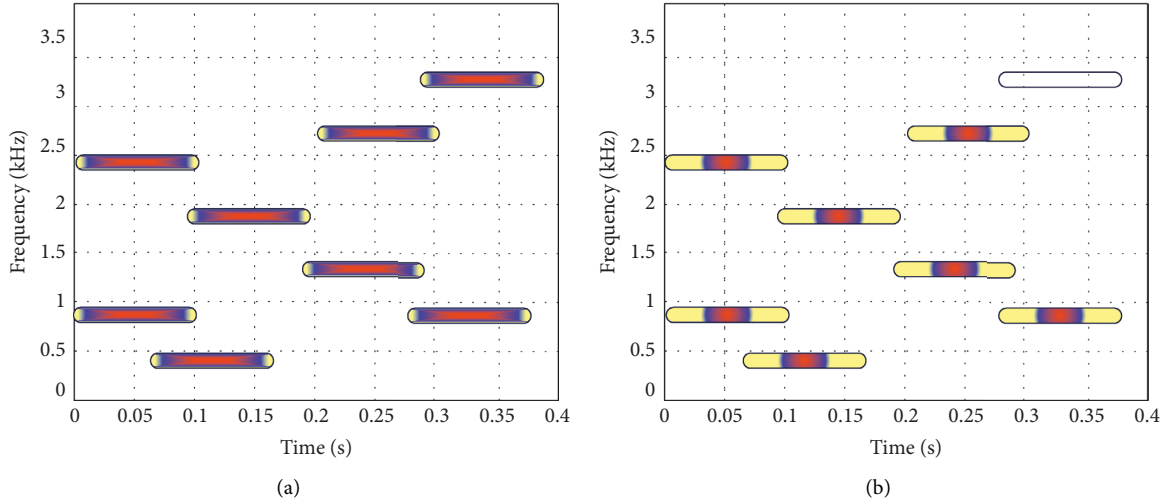


FIGURE 5: Two commonly used linear transformations. (a) Gabor transfer and (b) wavelet transform.

Among them, $\psi(t)$ represents the mother wavelet, a represents the scale, which controls the expansion and contraction of the wavelet function, and b represents the translation amount, which controls the movement on the time axis of the wavelet function. Figure 5 shows the analysis of trombone timbre frequency signal by Gabor transform and wavelet transform. The Gabor transform uses a Gaussian window with a window width of $N/(10 + 1)$, and the wavelet base used by the wavelet is a complex-valued Morlet wavelet.

It can be seen from Figure 5 that the Gabor transform is similar to STFT, that is, the resolution for each frequency is consistent. However, the wavelet transform has different resolutions for different frequencies, indicating that the resolution of the wavelet transform is adaptive. It shows that the wavelet transform has a good analysis effect for the signal of a single frequency, but for the multicomponent signal such as the trombone timbre frequency, the effect of the wavelet transform is not very good.

Compared with the linear time-frequency transformation, it does not have linear superposition, so after the transformation, there will be two results of self-term and

cross-term. The WVD transformation expression is given as follows:

$$\text{WVD}(t, f) = \int_{-\infty}^{\infty} s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) e^{j2\pi f\tau} d\tau. \quad (14)$$

It can be seen from Equation (14) that WVD does not need to select a window function similar to STFT transform to intercept the signal, and $s((t + \tau)/2)s^*((t - \tau)/2)$ can be regarded as the autocorrelation function of the signal. However, WVD is its Fourier transform with respect to τ , and the result obtained is a two-dimensional parameter of the time-frequency plane.

First, the WVD is processed by windowing in the time domain similar to STFT, which can remove the influence of the cross term, and then obtain the PWVD, and its expression is given as follows:

$$\text{PWVD}_x(t, f) = \int_{-\infty}^{+\infty} s^*\left(t - \frac{\tau}{2}\right) s\left(t + \frac{\tau}{2}\right) h(\tau) e^{-j2\pi f\tau} d\tau. \quad (15)$$

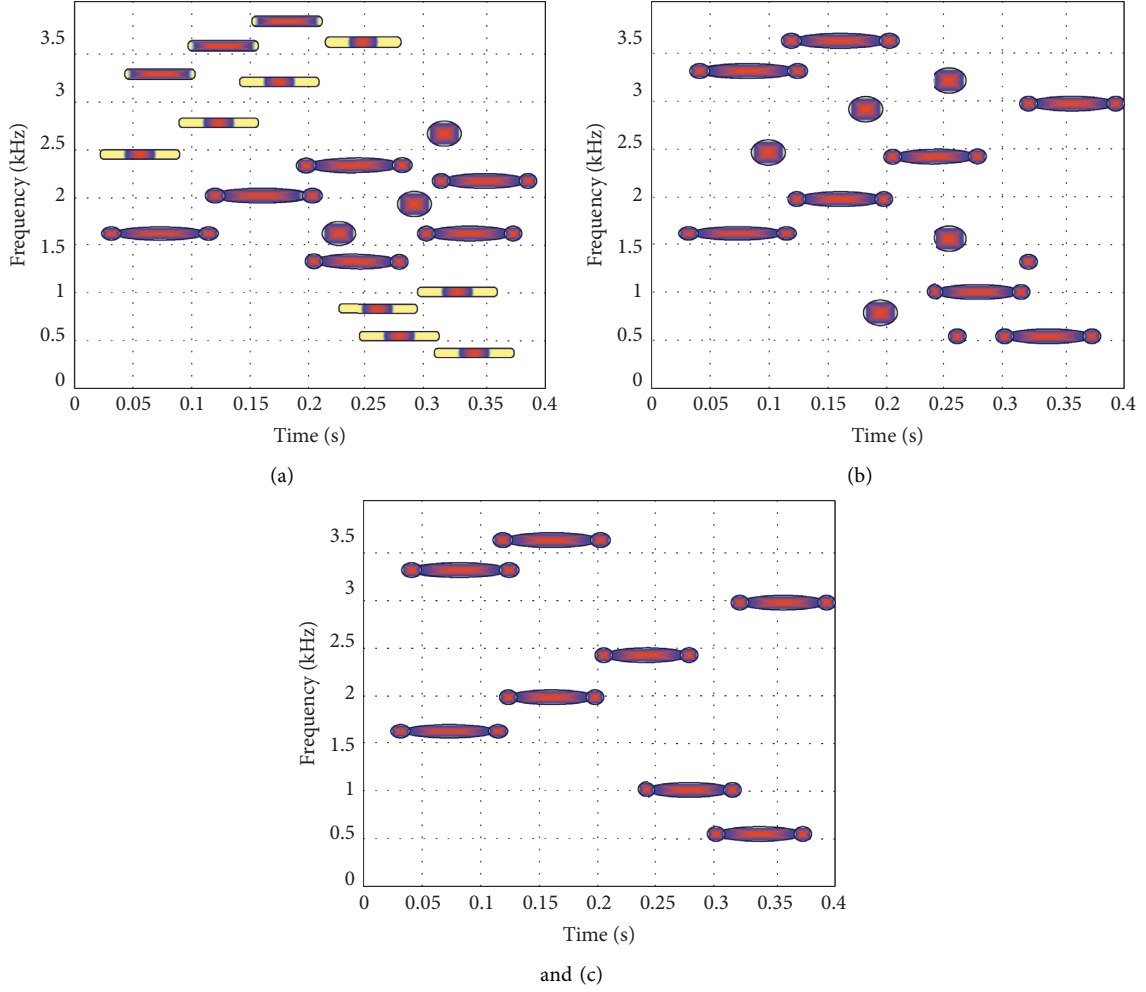


FIGURE 6: Time-frequency diagram of WVD and its improvement. (a) WVD, (b) PWVD, and (c) SPWVD.

The h (trombone timbre) in the formula is the added time window function, and its main function is to smooth the $s(t)$ in the time domain. PWVD reduces the impact of most cross-interference on the signal by sacrificing time-frequency focus.

In order to achieve the purpose of completely eliminating the cross term, based on the PWVD, a windowing operation is also performed in the frequency domain, which is SPWVD, and its expression is given as follows:

$$\text{SPWVD}(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s\left(t - \nu + \frac{\tau}{2}\right) s^*\left(t - \nu - \frac{\tau}{2}\right) h(t) g(\nu) e^{-j2\pi f\tau} d\nu d\tau. \quad (16)$$

In the formula, h (trombone timbre) represents the window function in the time domain, which mainly performs filtering in the time domain. g (trombone timbre) represents the window function in the frequency domain, and its function is to filter in the frequency domain. Compared with PWVD, because the window function is added in the frequency domain, the suppression effect of the cross term is better. However, secondary windowing makes SPWVD less time-frequency focused. Figure 6 is a time-frequency analysis diagram of WVD, PWVD, and SPWVD.

As can be seen from Figure 6(a), after the trombone timbre frequency signal undergoes WVD transformation,

the self-term and cross-term interference become indistinguishable, and even some cross-term energy is completely larger than the signal self-term energy. However, its time-frequency focus is indeed the best. Figure 6(b) shows the PWVD windowed in the time domain. It can be seen that it has a certain inhibitory effect on the cross term but does not completely eliminate the interference of the cross term, and the addition of the window reduces the time-frequency focus of the PWVD. Figure 6(c) shows the SPWVD with windows in both the time domain and the frequency domain. It can be seen that the interference of the cross term has been completely suppressed. However, the time-frequency focus

is also reduced, so the windowing operation is to suppress the cross-term problem between the signals at the expense of reducing the time-frequency focus.

The reason why the cross-term is generated, when WVD analyzes multiple signals, can be demonstrated by mathematical formulas.

For signal $s = s_1 + s_2$, there is

$$\begin{aligned}
 \text{WVD}_s &= \int_{-\infty}^{+\infty} s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &= \int_{-\infty}^{+\infty} (s_1 + s_2)\left(t + \frac{\tau}{2}\right) (s_1 + s_2)^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &= \int_{-\infty}^{+\infty} s_1\left(t + \frac{\tau}{2}\right) s_1^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &\quad + \int_{-\infty}^{+\infty} s_2\left(t + \frac{\tau}{2}\right) s_2^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &\quad + \int_{-\infty}^{+\infty} s_1\left(t + \frac{\tau}{2}\right) s_2^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &\quad + \int_{-\infty}^{+\infty} s_2\left(t + \frac{\tau}{2}\right) s_1^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \\
 &= \text{WVD}_{s_1} + \text{WVD}_{s_2} + 2\text{Re}[\text{WVD}_{s_1, s_2}].
 \end{aligned} \tag{17}$$

Among them, $2\text{Re}[\text{WVD}_{s_1, s_2}]$ is the introduced cross-interference term. Figure 6(a) can clearly see the cross-term between signals, while PWVD and SPWVD only filter out the cross-term part through the principle of time domain and frequency domain filtering. Therefore, the trombone timbre frequency signal is generally not analyzed by WVD.

In addition to the abovementioned secondary time-frequency analysis, spectrogram (SP) is also an important secondary time-frequency analysis method. The spectrogram is mainly obtained by the square of the STFT mode, and its expression is given as follows:

$$\begin{aligned}
 \text{SPEC}(t, f) &= |\text{STFT}(t, f)|^2 \\
 &= \int s(\tau) h^*(\tau - t) e^{-j2\pi f\tau} d\tau \\
 &\quad \int s^*(\tau) h(\tau - t) e^{-j2\pi f\tau} d\tau \\
 &= \iint W(\tau, \nu) W_h^*(\tau - t, \nu - f) d\tau d\nu.
 \end{aligned} \tag{18}$$

In the formula, the (τ, ν) operator represents the WVD transformation, so the spectrogram can be regarded as the two-dimensional convolution of the WVD of the signal and the WVD of the window function. The spectral operation is simple, so there are many practical applications. For multiple signals, such as $s = s_1 + s_2$, phase information is produced because the STFT result is a complex number. Its spectral expression is given as follows:

$$\begin{aligned}
 \text{SPEC}_s(t, f) &= |a\text{STF}_{s_1}(t, f) + b\text{STFT}_{s_2}(t, f)|^2 \\
 &= a^2 |\text{STFT}_{s_1}(t, f)|^2 + b^2 |\text{STFT}_{s_2}(t, f)|^2 \\
 &\quad + 2ab\text{STFT}_{s_1}(t, f)\text{STFT}_{s_2}^*(t, f) \cos(\phi_{s_1}(t, f) - \phi_{s_2}(t, f)).
 \end{aligned} \tag{19}$$

Among them, $\phi_{s_1}(t, f) = \arg(\text{STFT}_{s_1}(t, f))$, $\phi_{s_2}(t, f) = \arg(\text{STFT}_{s_2}(t, f))$. It can be seen that the spectrum does not satisfy the linear superposition, and there is a phase cross term of $\cos(\phi_{s_1}(t, f) - \phi_{s_2}(t, f))$. However, if the two signals do not overlap in the time and frequency domains, the following expressions are satisfied:

$$\begin{aligned}
 \text{SPEC}_s(t, f) &= |a\text{STF}_{s_1}(t, f) + b\text{STFT}_{s_2}(t, f)|^2 \\
 &= a^2 |\text{STFT}_{s_1}(t, f)|^2 + b^2 |\text{STFT}_{s_2}(t, f)|^2.
 \end{aligned} \tag{20}$$

Figure 7 is the spectrogram analysis of the trombone timbre frequency signal, in which the window function is selected as the hammering window, and the length of the window is $N/(10 + 1)$.

It can be seen from Figure 7 that there is no cross-term interference when the spectrogram analyzes a single trombone timbre frequency signal because the frequencies of the trombone timbre frequency signal in each time period do not coincide. The time-frequency focus of the spectrogram is also very high, and its computational complexity is smaller than that of SPWVD, so it is widely used in engineering practice. Of course, there are also analysis methods such as fourth-order spectrogram and eighth-order spectrogram, which will not be repeated here.

The received signal is processed by STFT, and then the time-frequency analysis diagram of the multitrombone timbre frequency signal can be obtained. Then, according to the signal-to-noise ratio, the corresponding cut-off value is set, and then the elements in the STFT matrix are comparatively cut off. Therefore, by truncating $\text{STFT}_s(t, f)$, $\text{STFT}'_s(t, f)$ is obtained, and its expression is given as follows:

$$\text{STFT}'_s(t, f) = \begin{cases} \text{STFT}_s(t, f), & \text{if } |\text{STFT}_s(t, f)| \geq \varepsilon, \\ 0, & \text{if } |\text{STFT}_s(t, f)| < \varepsilon. \end{cases} \tag{21}$$

In the formula, ε is the truncation threshold. If $\text{STFT}'_s(t, f)$ is discretized, $\text{STFT}'_s(n, m)$ can be obtained, where n is the sampling point on the time axis, the total number of sampling points is N , m is the sampling point on the frequency axis, and the total number of sampling points is M . Therefore, the threshold ε is defined as follows:

$$\varepsilon = \eta \times \text{mean}(\text{STFT}_s(n, m)) = \eta \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M \text{STFT}_s(n, m). \tag{22}$$

In the formula, η is the threshold factor.

The time-frequency analysis result $\text{STFT}'_s(t, f)$ after threshold truncation processing and $\text{WVD}_s(t, f)$ with good time-frequency focusing are processed by dot product,

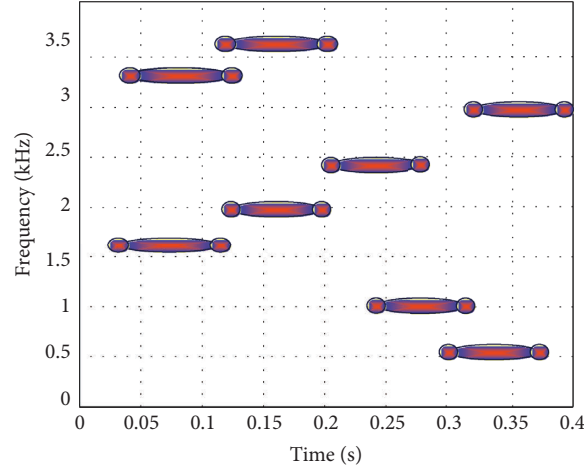


FIGURE 7: Spectrogram time-frequency analysis diagram.

namely, Hadamard product and the combined time-frequency distribution can be obtained as follows:

$$TF_s(t, f) = STFT'_s(t, f) \odot WVD_s(t, f). \quad (23)$$

Furthermore, good performance can be obtained by estimating the parameters of the combined time-frequency analysis signal. Taking the combination of STFT and WVD as an example, another combined time-frequency analysis is processed in the same way.

Figure 8 shows the time-frequency analysis of three common combinations of STFT-WVD, STFT-PWVD, and STFT-SPWVD and the time-frequency analysis diagram of STFT in the same noise environment. The time-frequency matrix transformed by the STFT is intercepted with a fixed threshold to make the time-frequency matrix show a distribution state of zero and nonzero values, so as to achieve the purpose of artificially weakening the noise.

It can be seen from the abovementioned figure that the time-frequency diagram of the combined time-frequency can suppress a certain amount of noise and is clearer and more robust than the single time-frequency analysis. Therefore, in the time-frequency analysis of multitrombone timbre frequencies, a combined time-frequency analysis may be an optimal solution. It can be observed that STFT-WVD and STFT-PWVD have little noise interference at multiple trombone timbre frequencies, so STFT-SPWVD time-frequency analysis is used in postprocessing signals.

The quality of the time-frequency analysis method is mainly measured from the time-frequency focusing and the cross-interference term. However, the level of time-frequency focus and the presence or absence of cross-terms can only be distinguished by the human eye, and a performance index based on information entropy to measure its time-frequency map is proposed. The specific steps are given as follows:

- (1) The algorithm uses the time-frequency analysis method to obtain the time-frequency analysis diagram for the signal, and then performs modulo processing on the elements in the obtained time-frequency matrix, and the modulo time-frequency matrix is expressed as $TF_x(t, f)$.

- (2) The algorithm finds the global maximum value $TF_{\max}(t, f)$ and the global minimum value $TF_{\min}(t, f)$ in $TF_x(t, f)$, and sets the step size of the amplitude interval as $\text{step} = \text{trombone timbre}/N$ and N represents the number of intervals to be divided.

$$\begin{aligned} & [TF_{\min}(t, f), TF_{\min}(t, f) + \text{step}), \\ & [TF_{\min}(t, f) + \text{step}, TF_{\min}(t, f) + 2 \text{ step}), \\ & \dots \\ & [TF_{\min}(t, f) + (N - 1)\text{step}, TF_{\max}(t, f)). \end{aligned} \quad (24)$$

The algorithm counts the amplitude values falling in these N intervals and obtains the frequency vector of amplitude values as $M = [m_1, m_2, \dots, m_N]$, and then divides it by the total number of elements in the time-frequency matrix to obtain N groups of probabilities $P = [p_1, p_2, \dots, p_N]$.

- (3) The algorithm calculates the information entropy of the time-frequency analysis graph with the following formula:

$$S = -K \sum_{n=1}^N p_n \ln p_n. \quad (25)$$

Through the obtained entropy value, the quality of various time-frequency analysis methods can be analyzed. The larger the entropy value is, the worse the time-frequency focusing is, and the worse the effect of suppressing cross-interference is. On the contrary, the smaller the entropy value is, the better the time-frequency focusing is, and the better the effect of suppressing cross-interference is.

3. Trombone Timbre Feature Extraction based on CNN Model

The general framework of trombone timbre feature recognition is shown in Figure 9. In the preprocessing stage, the musical tone samples are deaveraged and normalized to

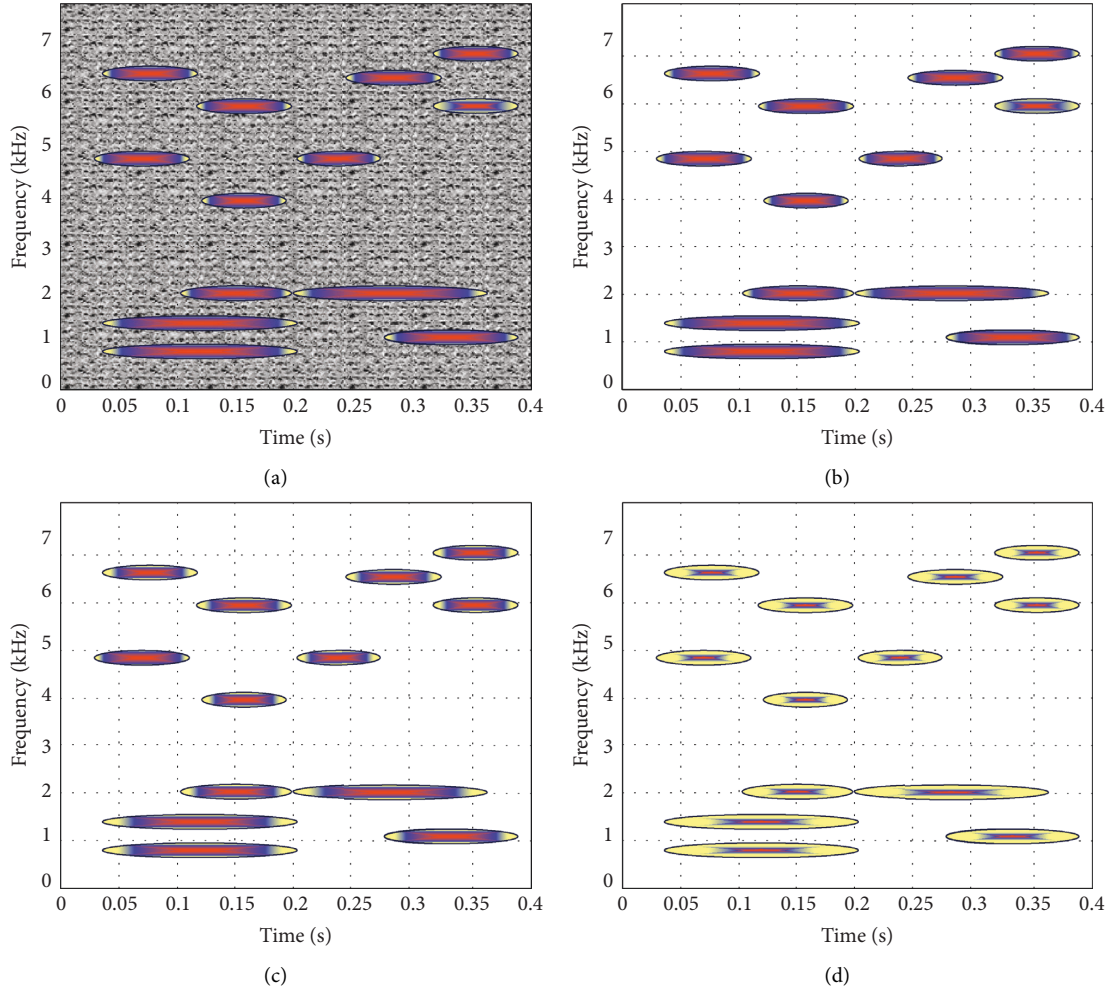


FIGURE 8: Time-frequency distribution of STFT and its combination. (a) STFT, (b) STFT-WVD, (c) STFT-PWVD, and (d) STFT-SPWVD.

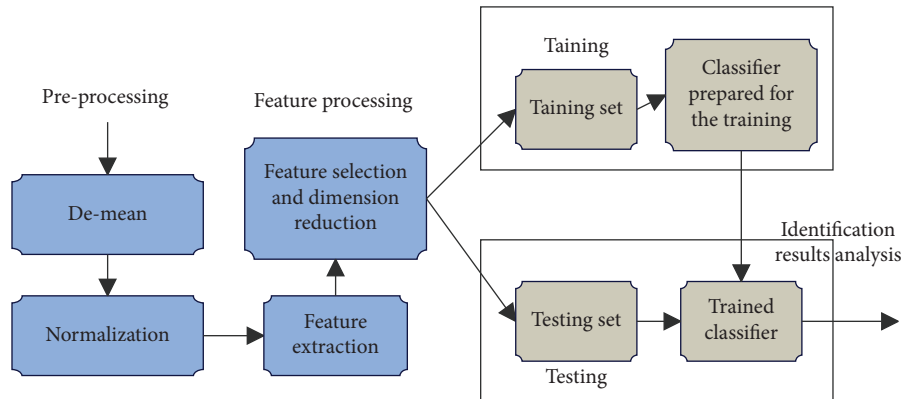


FIGURE 9: General framework of trombone timbre feature recognition.

avoid fluctuations in the mean and amplitude from affecting the stability of the final model. In the feature processing stage, timbre features, such as MFCC, are extracted.

In this paper, the feature recognition of the music performed by the trombone is carried out. Figure 10(a) and

Figure 10(b) are the loudness diagrams and the normalized energy changes and note onset diagrams of the trombone playing segment.

On the basis of the abovementioned research, the trombone timbre feature extraction method based on the

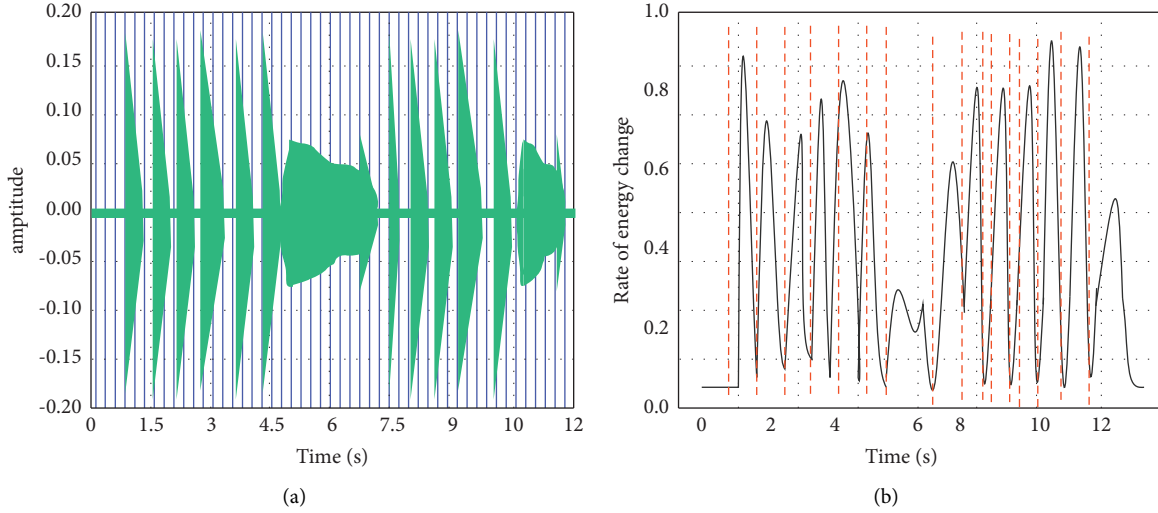


FIGURE 10: Recognition of musical features of trombone performance. (a) Performance section loudness and (b) schematic diagram of energy change and note onset.

TABLE 1: Accuracy of trombone timbre feature extraction based on the CNN model.

Num	Accuracy (%)
1	96.596
2	96.894
3	94.641
4	92.261
5	95.843
6	95.197
7	94.792
8	96.391
9	95.713
10	92.192
11	96.639
12	95.870
13	93.771
14	96.725
15	93.189
16	93.969
17	93.313
18	93.680
19	94.964
20	93.144
21	92.075
22	94.132
23	95.261
24	93.835
25	95.069
26	95.382
27	96.601
28	93.989
29	93.865
30	95.699
31	94.760
32	93.008

CNN model proposed in this paper is performed effectively, and the accuracy of the trombone timbre extraction is calculated, and the results shown in Table 1 are obtained.

It can be seen from the abovementioned research results that the trombone timbre feature extraction method based on the CNN model proposed in this paper can effectively identify the trombone timbre in a variety of musical performances.

4. Conclusion

The trombone, also known as the trombone and the telescopic horn, belongs to the brass playing instrument. It is the only musical instrument that has not been greatly improved in shape and structure since its origin. The trombone originated in BC and was first used in church and opera performances. In the nineteenth century, the trombone entered the symphony camp. Because of its unique timbre, it was mainly used in the performance of military bands, showing an impassioned, majestic, and powerful momentum. In addition, the trombone is also used in jazz performance, and the trombone is also known as the “king of jazz.” In this paper, the trombone timbre feature extraction model is constructed by combining the CNN model to improve the training and learning effect of the trombone. The simulation results show that the trombone timbre feature extraction method based on the CNN model proposed in this paper can effectively identify the trombone timbre in various musical performances.

Data Availability

The labeled dataset used to support the findings of this study is available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Shandong College of Arts.

References

- [1] X. Serra, "The computational study of a musical culture through its digital traces," *Acta Musicologica*, vol. 89, no. 1, pp. 24–44, 2017.
- [2] I. B. Gorbunova and N. N. Petrova, "Digital sets of instruments in the system of contemporary artistic education in music: socio-cultural aspect," *Journal of Critical Reviews*, vol. 7, no. 19, pp. 982–989, 2020.
- [3] E. Partesotti, A. Peñalba, and J. Manzolli, "Digital instruments and their uses in music therapy," *Nordic Journal of Music Therapy*, vol. 27, no. 5, pp. 399–418, 2018.
- [4] B. Babich, "Musical "covers" and the culture industry: from antiquity to the age of digital reproducibility," *Research in Phenomenology*, vol. 48, no. 3, pp. 385–407, 2018.
- [5] L. L. Gonçalves and F. L. Schiavoni, "Creating digital musical instruments with libmosaic-sound and mosaiccode," *Revista de Informática Teórica e Aplicada*, vol. 27, no. 4, pp. 95–107, 2020.
- [6] I. B. Gorbunova, "Music computer technologies in the perspective of digital humanities, arts, and researches," *Opción*, vol. 35, no. 24, pp. 360–375, 2019.
- [7] A. Dickens, C. Greenhalgh, and B. Koleva, "Facilitating accessibility in performance: participatory design for digital musical instruments," *Journal of the Audio Engineering Society*, vol. 66, no. 4, pp. 211–219, 2018.
- [8] O. Y. Vereshchahina-Biliavska, O. V. Cherkashyna, Y. O. Moskvichova, O. M. Yakymchuk, and O. V. Lys, "Anthropological view on the history of musical art," *Linguistics and Culture Review*, vol. 5, no. S2, pp. 108–120, 2021.
- [9] A. C. Tabuena, "Chord-interval, direct-familiarization, musical instrument digital interface, circle of fifths, and functions as basic piano accompaniment transposition techniques," *International Journal of Research Publications*, vol. 66, no. 1, pp. 1–11, 2020.
- [10] L. Turchet and M. Barthet, "An ubiquitous smart guitar system for collaborative musical practice," *Journal of New Music Research*, vol. 48, no. 4, pp. 352–365, 2019.
- [11] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, and S. Jänicke, "A survey on visualizations for musical data," *Computer Graphics Forum*, vol. 39, no. 6, pp. 82–110, 2020.
- [12] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F. R. Stöter, "Musical source separation: an introduction," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 31–40, 2019.
- [13] T. Magnusson, "The migration of musical instruments: on the socio-technological conditions of musical evolution," *Journal of New Music Research*, vol. 50, no. 2, pp. 175–183, 2021.
- [14] I. B. Gorbunova and N. N. Petrova, "Music computer technologies, supply chain strategy and transformation processes in socio-cultural paradigm of performing art: using digital button accordion," *International Journal of Supply Chain Management*, vol. 8, no. 6, pp. 436–445, 2019.
- [15] J. A. Anaya Amarillas, "David Andrés Martín marketing musical: música, digital knowledge sharing in the age of music, industry and promotion BY-NC-ND 3.0, eBook," *Inter Disciplina*, vol. 9, no. 25, pp. 333–335, 2021.
- [16] G. Scavone and J. O. Smith, "A landmark article on nonlinear time-domain modeling in musical acoustics," *Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. R3–R4, 2021.
- [17] L. Turchet, T. West, and M. M. Wanderley, "Touching the audience: musical haptic wearables for augmented and participatory live music performances," *Personal and Ubiquitous Computing*, vol. 25, no. 4, pp. 749–769, 2021.

Copyright of Journal of Robotics is the property of Hindawi Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.