Evaluations of Clarinet Sound Quality and Register Characteristics

Jiasheng Xu *

Faculty of Arts and Sciences, Beijing Normal University, Beijing, China

* Corresponding Author Email: 202011059057@mail.bnu.edu.cn

Abstract. The clarinet is a commonly used woodwind instrument that has a unique sound quality, making it a prominent instrument in solo performances and orchestral music. Therefore, sound quality plays an important role in the perception of clarinet sound. This study provides a review of the relevant research in the field of clarinet sound analysis and synthesis, and identifies many valuable references from previous studies. In studying the noise component of clarinet sound, this study recorded audio samples of E3, E4, and E5 and preprocessed them by slicing and performing frequency domain transformations. Using 13 representative spectral features, this paper extracted features from the spectral signals and constructed a classification model of sound quality based on the SVM algorithm of machine learning. Finally, this study analyzed the shortcomings of the SVM algorithm in high-pitch analysis and proposed future research directions of combining the SVM algorithm with different sound characteristics of clarinet registers and spectral changes throughout the entire process of playing. These results shed light on guiding further exploration of the timbre of the clarinet.

Keywords: Clarinet; sound analysis; SVM.

1. Introduction

Computer music is a unique and fascinating blend of technology and artistic expression that dates to the 1960s [1]. It has captivated music enthusiasts since its inception and continues to evolve with advancements in computer technology and software development. One area of focus in computer music is timbre analysis, which plays a critical role in understanding and creating the distinct sounds of different musical instruments [2]. The clarinet, a popular woodwind instrument, has a distinct and recognizable timbre that sets it apart from other instruments [3]. Given its unique acoustic structure and the various ways in which it can be played, analyzing, and researching the clarinet's timbre is particularly important in understanding and appreciating its sound.

Moorer introduces the application of modern digital signal processing technology in the field of music, providing unprecedented freedom and precise control for composers and musicians [4]. The topics covered include natural sound analysis, speech and vocoder technology, artificial reverb methods, and efficient synthesis of discrete summation formulas. Lazzarini mainly introduces the development history and technical characteristics of computer music programming systems, tracing the origin and development of computer music programming from early offline synthesis programs to modern real-time interactive systems [5]. When discussing computer music programming systems, scholars also involved some audio processing-related technologies, e.g., digital signal processing, filter design, and sampling rate conversion. Loy and Abbott introduce some common music programming languages (e.g., MUSIC V, Csound, CMusic, and LISP), which typically have powerful generative capabilities and implement special music data types and operations, as well as methods for representing common forms in music such as time flow, recursion, and parallelism [6]. Beauchamp and Horner state that spectral modeling and timbre blending programs can help visualize the time-varying timbre of sound, discover good synthesis parameter settings that match specific instruments, and assist in sound synthesis [7]. This research used traditional methods such as frequency modulation, wavetable interpolation, and nonlinear filtering synthesis to solve the problems of spectral modeling and timbre blending in computer music. In their analysis of clarinet reeds and sound quality, Wang, Guan, and Du employed a new method based on harmonic structure

and energy distribution to assess clarinet reed quality [8]. The researchers discovered that reed quality is closely linked to even-numbered harmonics by separating the reed and clarinet mass. They then created a feature set that includes the even harmonics envelope and harmonic energy distribution. The team recorded annotated clarinet audio data from three levels of performers and employed machine learning support vector machines to classify the sound quality. In another study of clarinet equivalent tube length and sound, Dickens, France, and Smith used a web-based database to investigate the acoustic specifics of the clarinet, which included standard fingerings and other fingerings [9]. The clarinet has a cylindrical tube with one end acoustically closed and the other open, making it an ideal example of a closed-open tube. By comparing measurements of cylindrical bodies of equivalent acoustic length and those of clarinets, several phenomena can be clarified. Barthet, Guillemain, and Kronland-Martinet studied the correlation between control gestures, produced timbres, and their perceptual representations of the clarinet [10]. This research provides significant support for synthesis and control, music analysis and perception, and music information retrieval. By using a physics-based model to alter the primary control parameters of the model (related to blowing pressure and lip pressure on the reed), they generated synthesized clarinet tones and obtained the low-dimensional configuration that best represents the difference rating through multidimensional scaling and hierarchical clustering of the sound data collected in the experiment. They discovered that the even-odd ratio is a useful indicator for predicting the unique vibrating reed situation of the clarinet by comparing the sounds of natural and synthesized clarinets.

The literature above not only presents the history of computer music development but also covers research on various aspects of clarinet acoustics and performance, providing important scientific support and research methods for a deeper understanding of clarinet timbre analysis and its sources. It also indicates that there are already many mature analysis models and tools in the field of computer science, and there have been many rich studies on clarinet sound analysis and its spectral domain. However, despite the numerous and rich studies on clarinet timbre, there is still a research gap in the specialized study of noise generated by factors such as the reed, mouth shape, and clarinet structure. What one perceives as "unfocused" clarinet sound is often caused by the excessive noise component. Therefore, the study of timbre will mainly focus on this aspect. In this paper, the researchers start by identifying potential areas of study related to clarinet sounds. They proceed to record a range of sounds from the clarinet, including both obvious and subtle noises, and utilize Nyquist to take samples. Afterward, they analyze the sound frequency using MATLAB and apply machine learning techniques based on SVM to categorize the various sound types. Ultimately, they compare and evaluate the classification results of the sounds.

2. Data & Methodology

2.1. Data

The audio dataset used in this study was recorded using the professional portable recording device AKG-LYRA to ensure the recording quality and avoid external noise interference. The recordings were conducted in a quiet music rehearsal room late at night. At a fixed distance of 2 meters, long sound samples of clarinet playing E3, E4, and E5 were recorded for one hour each, using the same instrument played by both a skilled player and a beginner with obvious air flow noise problems. The recorded audio was then labeled and classified, with the audio played by the skilled player labeled as the first category and the beginner's audio as the second category.

To ensure consistency in the analyzed samples, the long sounds needed to be processed first. As each long sound is played with a gradual increase and decrease in volume, the sound is not linearly changing during this process. Therefore, it is necessary to trim the sound during the long sound. To eliminate interference from the beginning and end of the long sound, this study selected the middle section of the sound for analysis and converted the sound to mono. For each long sound, one second of audio is sufficient to demonstrate its sound characteristics. Therefore, for each long sound, the start time and duration were adjusted to trim the audio, and 60 one-second long sound samples were

trimmed for each pitch to obtain the time domain data of the samples. To better visualize the differences between the samples, it is necessary to perform frequency domain transformation on each sample. Due to the large number of samples and the need to use machine learning tools in subsequent processes, MATLAB was used as a tool to transform the sound slices into frequency domain signals in matrix form.

The Discrete Fourier Transform (DFT) is the most widely used frequency domain transformation tool:

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right) \quad (1)$$

In general computer applications, the Fast Fourier Transform (FFT) is used instead. It converts a finite-length time-domain sequence into a frequency-domain and discretized finite-length sequence through the above DFT. Based on symmetry and repetition properties, FFT reduces the computational complexity from $O(n^2)$ to O(n log n), greatly improving the computational speed. Fig. 1 shows the FFT transformation result of sample E3-1 (the first sample in E3) as an example.

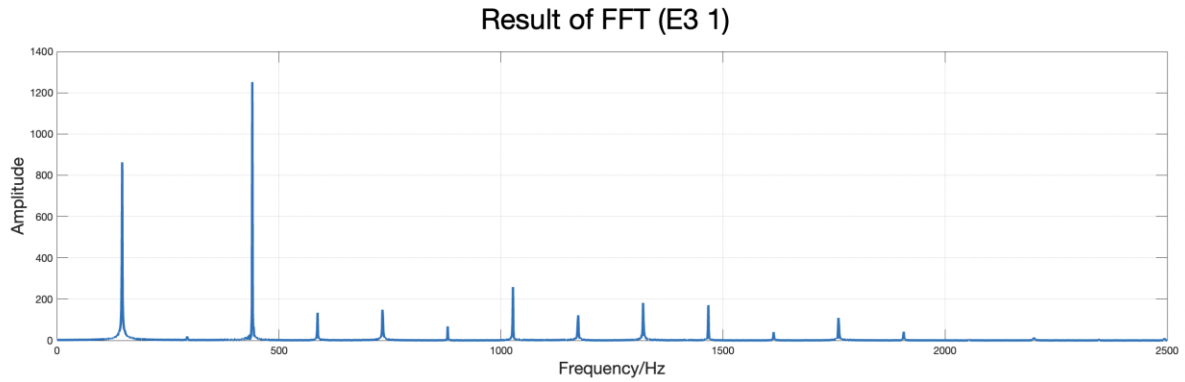


Fig 1. An Example of Frequency Domain Signal Processing.

2.2. Models & Evaluations

After performing the Fast Fourier Transform, the frequency domain data for all samples of E3, E4, and E5 were obtained. According to the sampling theorem, the data was transformed into three matrices of 60 x 22050. To further explore the data, this study used feature extraction to select 13 representative audio features and transformed these three frequency domain matrices into feature matrices of audio samples, as shown in Table 1.

Table 1. Features.

RMS	Crest Factor	Shape Factor	Impulse Factor	Peak to Peak
Margin Factor	Sum	Mean	Std	Min
Skewness	Kurtosis	Max		

When finish obtaining the feature matrix, this study used SVM to perform machine learning and classification based on the information with the aim of automatically identifying noise components in the performance through the training set. SVM is a type of generalized linear classifier that performs binary classification on data in a supervised learning manner, and its decision boundary is the maximum margin hyperplane obtained by solving the learning samples.

MATLAB's machine learning toolbox was used in this study to perform SVM classification on the data. The two-dimensional corresponds to two feature vectors, and when there are 13 feature vectors, the maximum margin line for the 13 dimensions can be solved according to the two-dimensional method. However, not all points are linearly separable, and this study used a kernel function to expand the data dimension and build the decision model. The following are the algorithm steps in the SVM classification process:

(1) Clearing environment variables, turning off warning messages, closing open figure windows, clearing variables, and clearing the command window.

(2) Reading matrix data from an Excel file and storing it in the "res" variable.

(3) Randomly dividing the data into training and test sets, where 40 samples are in the training set and 20 are in the test set.

(4) Normalizing the data using mapminmax function.

(5) Transposing the training and test data to fit the SVM model.

(6) Creating an SVM model using the svmtrain function with a radial basis function kernel and specified penalty factor and gamma values.

(7) Testing the model's performance on the training and test data using the svmpredict function and calculating the classification error.

(8) Sorting the predicted results and the actual results for both the training and test sets.

(9) Plotting the comparison between the actual and predicted values for the training and test sets.

(10) Creating confusion matrices for the training and test sets using the confusion chart function.

The advantage of the above algorithm is that it can judge the training effect of the model by comparing the differences between the test set and the training set, and the confidence matrix can be used to determine its reliability.

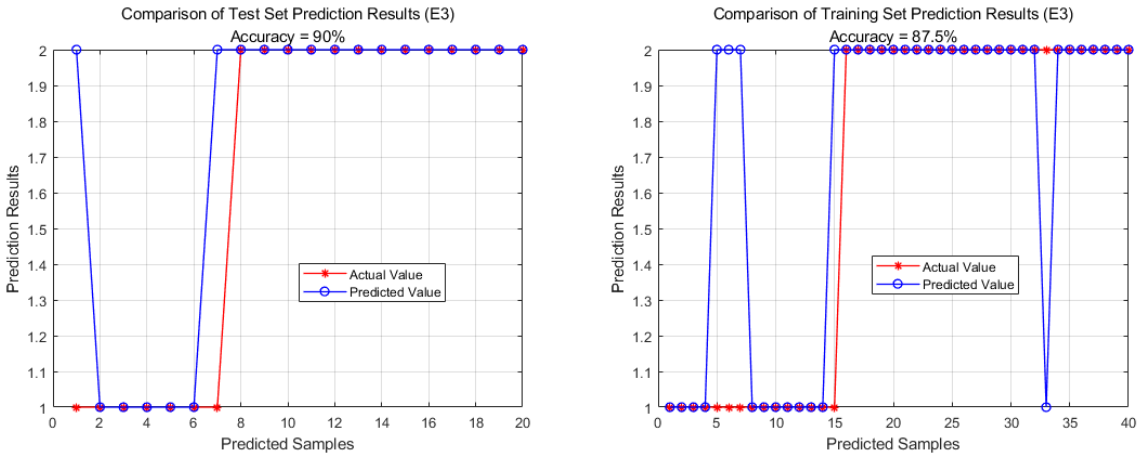


Fig 2. SVM Classification Results for E3.

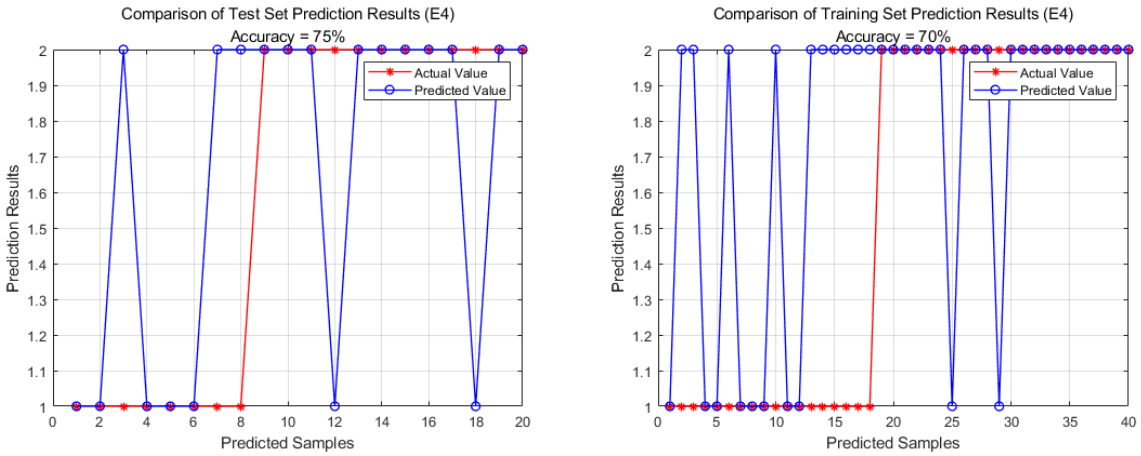


Fig 3. SVM Classification Results for E4.

3. Results & Discussion

In practical applications, this study selected 40 samples as the training set and 20 samples as the test set. All data was normalized and penalty factors as well as radial basis function parameters were constructed for simulation testing. In the data processing process, this study had already classified the

sound quality into good and ordinary sound based on their characteristics, so only the data from E3, E4, and E5 needed to be imported. When plotting the graphs, this study used red dots to represent the actual values and blue dots to represent the predicted values, comparing the fitting degree of the test set and the training set. The Fig. 2, Fig. 3, and Fig. 4 are the results of the training and testing.

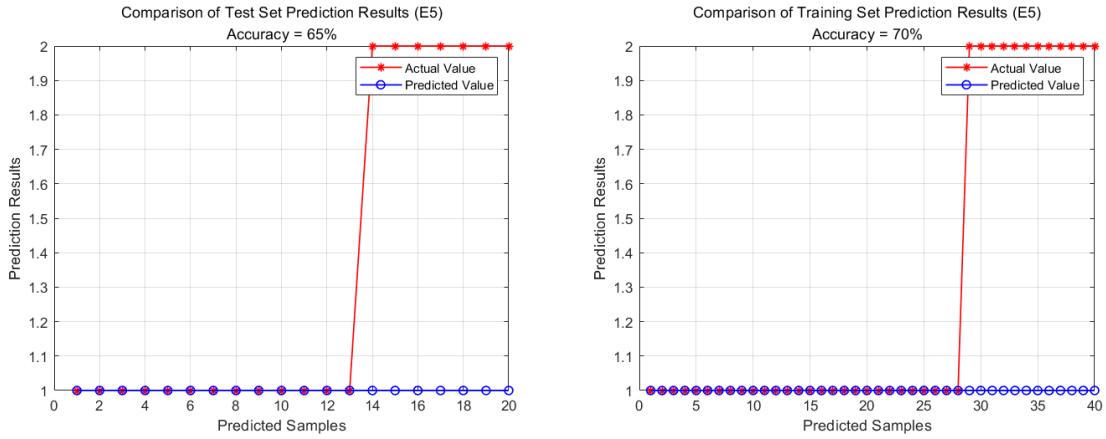


Fig 4. SVM Classification Results for E5.

According to the results, it can be found that the low frequency E3 part can distinguish these two categories very well, with an accuracy of around 90% for both the training set and the testing set. In the mid-frequency E4, the fitting effect has dropped to some extent, with fitting accuracy of 75% and 70% respectively. The main reason for the decrease in the fitting accuracy in the training set is that it considers many good sounds as bad sounds. For the E5 that uses overtones, all sounds are judged as good sounds, and the training effect is very poor in both the training set and the testing set.

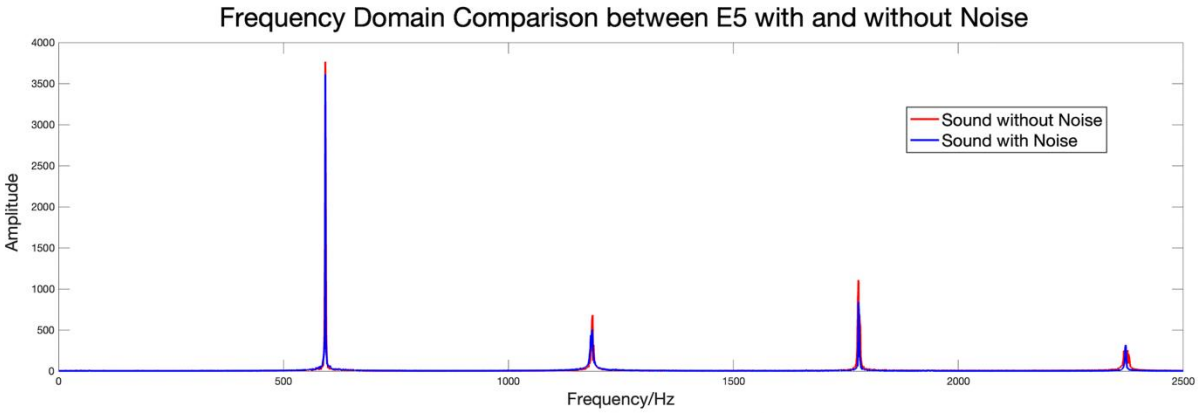


Fig 5. Frequency Domain Comparison of E5.

To determine the reason why the sound cannot be distinguished, all E5 data in this study were extracted and their mean data for good E5 sounds and bad E5 sounds were obtained according to their original classification. They were then imported into MATLAB for frequency domain analysis, and their frequency spectrum is shown in Fig 5. Through the comparison and analysis of the frequency spectrum, it can be found that the noise component is not significant and does not affect the features such as skewness and peak value that one has extracted. Specifically, for the frequency spectrum itself, good performance in playing produces fuller and louder sound at lower frequencies, and the adjacent sound volume is higher around specific frequency of the harmonic. In contrast, the sound with obvious noise has a lower volume at lower frequencies, and the sound is louder than that of the good performance in playing at the harmonic frequencies in the high-frequency range. However, these features cannot be reflected in the feature extraction mentioned above, which is also the reason why this model cannot classify the high-frequency range well.

4. Limitations & Prospects

The fitting performance of this model for the high-pitched sound range, especially the range that uses harmonic keys, is not satisfactory, and it is necessary to try selecting other more representative features. Based on the spectrum analysis of the sound in E5, the selected feature indicators should pay attention to the differences between the high and low frequency parts. In particular, for long sounds with noise, their noise components often manifest as other noise in the high frequency part.

In addition to the differences between sound ranges, there are also certain differences between each note within the same sound range. Many people who practice playing the clarinet, even if they do not have absolute pitch, can often judge which note they are playing based on changes in the timbre of the clarinet's sound. Therefore, when using other long sounds, such as F4 or F5, other issues may arise, and the experimenter may need to further optimize the selected features.

In the future, it is believed that constructing a complete model that combines the characteristics of different sounds will be a very meaningful task, in addition to optimizing the model itself. Furthermore, since the subsequent research also involves detailed study of the characteristics of each note, it is foreseeable to construct a more realistic and high-fidelity clarinet sound synthesizer model based on that. Moreover, this study is inspired by Professor Dannenberg's research on trumpet sound synthesis, which considers factors such as melodic contour, articulation, and dynamics, and uses Spectral Interpolation Synthesis to synthesize realistic trumpet performances [11]. In subsequent research, it should not be limited to the steady-state sound characteristics, as the sound onset and decay could also be important discriminative indicators.

5. Conclusion

In summary, this study investigated the sound characteristics of the clarinet based on spectral features, focusing on the differences in long tones during the stable phase of the sound and an in-depth analysis of the noise component. According to the analysis, SVM can be used to distinguish between clarinet long tones with and without noticeable noise, especially in the low-pitched sound range. Additionally, it is found that the specific cause of the noise component is the insufficient fullness of the low-frequency part of the sound and the excessive volume of the high-frequency harmonic part. Further study ought to improve the selected feature indicators, increase the fitting degree for analyzing high-pitched clarinet sounds, and further analyze long tones at other pitches. Overall, the research results can serve as an important guide for beginners and for selecting instruments.

References

[1] Chadabe J. Remarks on computer music culture[J]. Computer Music Journal, 2000, 24(4): 9-11.

[2] Pino D. The clarinet and clarinet playing[M]. Courier Corporation, 1998.

[3] Мякотин Е В. Timbre as a semiotic system. On the problem of structural analysis of music timbre[J]. PHILHARMONICA. International Music Journal, 2020 (6): 61-66.

[4] Moorer J A. Signal processing aspects of computer music: A survey. Proceedings of the IEEE, 1977, 65(8): 1108-1137.

[5] Lazzarini V. The development of computer music programming systems. Journal of New Music Research, 2013, 42(1): 97-110.

[6] Loy G, Abbott C. Programming languages for computer music synthesis, performance, and composition. ACM Computing Surveys (CSUR), 1985, 17(2): 235-265.

[7] Beauchamp J, Horner A. Spectral modelling and timbre hybridisation programs for computer music. Organised Sound, 1997, 2(3): 253-258.

[8] Wang Y, Guan X, Du Y, et al. Harmonics Based Representation in Clarinet Tone Quality Evaluation. IEEE, 2020.

[9] Dickens P, France R, Smith J, et al. Clarinet acoustics: introducing a compendium of impedance and sound spectra. Acoustics Australia, 2007, 35(1): 17.

[10] Barthet M, Guillemain P, Kronland-Martinet R, et al. From clarinet control to timbre perception. Acta Acustica united with Acustica, 2010, 96(4): 678-689.

[11] Dannenberg R D, Pellerin H. Derenyi,"A study of trumpet envelopes". Proceedings of the International Computer Music Conference (ICMC). 1998.