DOI: 10.23977/jaip.2024.070202
ISSN 2371-8412 Vol. 7 Num. 2

Exploration on Classification of Vocal Music Theme Based on Intelligent Multi Image Feature Fusion

Yue Zhu

Xi'an Shiyou University, Xi'an, Shaanxi, 710065, China

Keywords: Artificial Intelligence, Multi-image Feature Fusion, Classification of Vocal Music, Main Melody

Abstract: Music is a form of artistic expression, which can cultivate sentiment. Music can also be regarded as a language. Listening to music can give play to imagination through hearing, so as to resonate with the expression ideas of music creators. Music has no boundaries and is a cultural heritage. Therefore, when listening to music, music itself is not the purpose of listening, but the meaning behind music is the focus of feeling. People can communicate and express their feelings through music. The scope of music is very broad. Vocal music, as a kind of music, can be seen as an art combined with language. People's understanding of the classification of the main melody of vocal music is very simple. The music culture is broad and profound. The clear classification of the main melody of vocal music is of great significance to deepen the understanding of music. Therefore, this paper proposed an intelligent multi image feature fusion classification of vocal music theme. Through the experiment on the classification of vocal music theme by the artificial intelligence multi image feature fusion model, the data obtained showed that the number of single part music included in the 60 music classification was 17; the number of polyphonic music was 22, and the number of theme music was 21. It was consistent with the judgment results of two professional music teachers and student B, which indicated that intelligent multi image feature fusion can improve the accuracy of music classification. This study provided a reference value for the role of intelligent multi image feature fusion in the classification of vocal theme, and provided a future direction for the development of vocal theme classification.

1. Introduction

Music has been recorded in ancient Chinese books. Zhou Rites, Land Officials and Drum People: “The drum people teach the sound of six drums and four golden drums, make vocal music, harmonize with the army and serve in the field.” Li Fu of the Tang Dynasty once said in “The Mysterious Records of Qilin Guests”: “Singing phoenix, dancing phoenix and all kinds of vocal music are unheard of.” Music is not a collection of sounds, so not all sounds can be called music. Music is a collection of regular melodies used to express emotions. It can be said that music is a language with artistic expression. However, modern music generally refers to singing, and music accompaniment is optional. The singing throat is the main output form, which is different from instrumental music [1-2]. Since singing throat is involved, voice is the main singing form of music.

In other words, vocal music actually belongs to music. As a combination of music and literary language, vocal music plays an important role in beautifying Chinese art and enhancing artistic enjoyment. Music is invisible and intangible, and the classification of the main melody of vocal music cannot be described with words [3-4]. In order to carry forward the charm of vocal music and make the classification of vocal music theme more clear and distinctive, this paper looked for inspiration from intelligent multi image feature fusion based on the research of vocal music, and conducted relevant contrast experiments. It is hoped that the intelligent multi image feature fusion can provide inspiration for the classification of vocal music theme, so as to improve the classification of vocal music theme and promote the development of music art [5-6].

2. Exploration on the Classification of Vocal Theme

2.1 Music Overview

2.1.1 Music Concept

Music is an artistic expression form composed of regular and rhythmic melodies. There is no detailed explanation of the concept of music, and the understanding of music varies from person to person. Therefore, the connotation of music in different people's hearts is different, just as music itself can only be understood but not explained. Music perception is a positive listening behavior, which provides an irresistible cognition [7]. There are many functions that music brings to human perception. Different types of music bring people different perceptions, and the generation of these perceptions is beyond subjective control.

The style and characteristics of music referred to by the type of music can also be called the style of music. The unique characteristics of music as a whole are called the style of music. Through the music style, people can appreciate the ideological realm that the composer wants to express. Through music, people can have empathy and communication across borders. There are generally 7 types of music, which are shown in Figure 1.

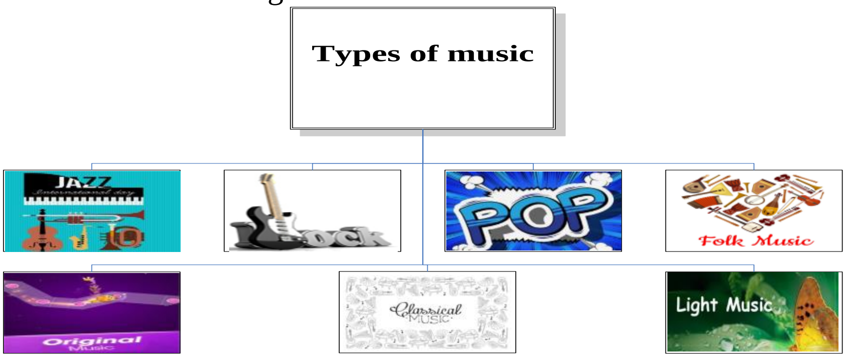

Figure 1: Music type diagram

This is just a rough division of music types. The melody composition of music is complex and diverse, and even can be divided into dozens of types in detail. Music can also be divided into two types: vocal music and instrumental music.

2.1.2 Vocal Concept

Vocal music refers to the musical form with human voice as the main form of expression, which can be accompanied by music or without music. As each person's timbre is different, vocal music can be divided into six categories according to timbre, which are shown in Figure 2.

Alto
Tenor
Mezzo-soprano
Baritone
Soprano
Bass

Figure 2: Type diagram of timbre

Music can be divided into different styles according to singing styles, and vocal music can also be divided into three types according to different singing modes:

① Bel canto

Bel canto originated in Italy in the middle of the 17th century, and its characteristics are crisp and euphemistic tone, beautiful and smooth tone. With a long history and centuries of development, Bel canto has attracted many experts and scholars to study it. The theory and method of Bel canto are also the most systematic and complete of the three singing modes. The vocalization of this singing method mainly depends on the basic laws of human physiological function, and follows the natural breathing mode and scientific vocalization way, so it can reduce the impact and burden of singing for a long time on the vocal cords, thus improving the service life of the vocal cords. It is well known that singers do not eat too stimulating food in their daily life, so as not to affect their vocal cords. They also pay great attention to the maintenance and care of their vocal cords at ordinary times. Bel canto singing can effectively protect their vocal cords. Many singers who are nearly 70 or 80 years old can certainly show their voice.

Generally, Bel canto singing can be divided into vertical singing and horizontal singing. Vertical singing refers to opening the mouth up and down to form an “O” shape when singing songs. The throat is also opened up and down, and the voice is made by using the vocal cords to stand up. Therefore, the voice produced by this singing method would give people a sense of surging and vigorous. And horizontal singing, as the name implies, is to open the mouth horizontally, exposing the teeth as if laughing. The throat is opened back and forth, and the tension of vocal cords is used for vocal singing. No matter horizontal singing or vertical singing, as long as it is properly used, it can bring the singing level to the best level.

Although bel canto originated in Italy, the style of bel canto can be divided into two schools. One is Italian school, the other is Russian school. China basically adopts Russian bel canto. Its characteristic is that when the singing voice is in the middle low voice, the sound area is in front; when the high voice is in the high voice, the sound area suddenly turns back; the singing method is horizontal singing, so the timbre is powerful.

② National singing

National singing represents the general name of folk singing with Chinese characteristics, which takes singing folk songs with national characteristics as its main form. China has 56 ethnic group and is a multi-ethnic country. Therefore, there are certain differences in the ethnic customs and emotions displayed by different regions, so the styles of national singing are also different. Therefore, before singing folk songs, understanding the characteristics of folk songs helps to strengthen the emotional understanding of folk songs.

Due to the large number of ethnic groups, the variety of Chinese folk songs is also quite rich and has a long history. In the evolution of China for five thousand years, folk songs with their own characteristics have been formed for all ethnic group. Han people’s minor in the south and folk songs in the north have the greatest influence in folk songs, and they continue to develop as the mainstream of Chinese folk songs. The southern minor is ethereal and melodious, with soft melody, and gives people a sense of calm, slow voice and slow tune. In terms of singing skills, national singing emphasizes “correct pronunciation” and attaches great importance to articulation clarity and rhythmic regularity.

③ Pop singing

Popular singing refers to the singing in the popular sense, which generally refers to the singing of popular songs. The concept of pop songs is not correct, because pop songs are different from other songs. Therefore, popular singing can also be called “commercial song singing”.

The popular singing method is easy to understand, popular, and basically does not need too many singing skills. Its simple singing characteristics have made it develop rapidly, and its influence even exceeds Bel canto and national singing methods. As pop singing has no fixed singing skills and patterns, and emphasizes randomness, “everyone can sing”. Popular singing also does not need a special theater or stage to perform, but can be natural humming and improvisation. Therefore, popular singing also has a mass character.

The early development of popular singing has not formed its own style, so it mainly depends on imitation, and then gradually mature into its own style. The pop singing method emphasizes the randomness of singing, which starts from the feeling and sings whenever people want. Therefore, more attention is paid to the “feeling” when singing. To become an excellent pop singer, it is not enough to have a good voice, but also needs strong imitation and excellent musical sense.

Although pop singing emphasizes the randomness of singing, it does not mean that people can sing randomly when singing songs. It also has the characteristics of pursuing individuality and characteristics. Therefore, pop singing can also be classified into different singing methods such as “Qi Sheng” and “Shouting”. The singing method of “Qi Sheng” needs to relax the throat, which basically maintains a state of inspiration when singing. The throat cannot be tightly closed when singing, but needs to be relaxed and in a state of air leakage. At this time, the voice is very ethereal. The singing principle of “shouting” depends on the roaring singing of the voice, which makes the voice strong and powerful. This singing method has an impact on the life of the vocal cords.

2.1.3 Theme Concept

Generally speaking, the theme is the theme, just like the central idea that the film and television works or literary works want to express. Usually there would be a climax in the performance of a piece of music. Through the performance of this part, the emotion of the composer or performer would be vividly displayed. It is usually expressed as a change or repetition of a musical rhythm or phrase.

2.2 Overview of Intelligent Multi Image Feature Fusion

Multi image feature fusion refers to the process of obtaining multiple information data with the same target through the recognition of multiple sensors, through computer, image processing and other technologies, and extracting the maximum information according to the unique features of the information data, so as to finally fuse into a high-quality image. Multi image feature fusion can effectively improve the contrast, clarity and color fidelity of low illumination images, and reduce the negative impact of image degradation [8].

Its advantage is that it can comprehensively process information data according to multi-source information, thus effectively improving the utilization and reliability of data. Multi image feature fusion has been widely used in many fields. Based on this advantage, this paper described the intelligent multi image feature fusion model. With the rapid development of artificial intelligence, multi image feature fusion has already been integrated with artificial intelligence. The development of the model has also evolved from the early immature stage to the current hybrid AI multi image feature fusion model, which is shown in Figure 3.

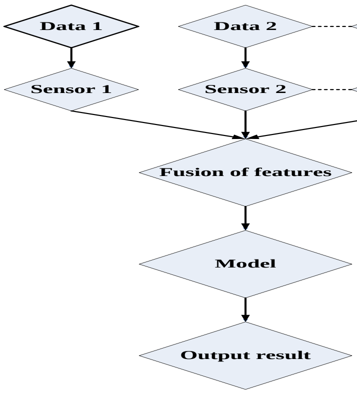

Figure 3: Multi image feature fusion model based on hybrid artificial intelligence

In the process of fusion transmission, data would inevitably be affected by external factors, such as time noise, spatial noise, etc. Therefore, different de-noising methods are required for the sensing process of the sensor, and the commonly used de-noising methods are smoothing and filtering. This paper adopted Wiener filtering and median filtering.

When the data signal is stable, wiener filtering works very well regardless of whether the data is discrete or continuous. The principle of least squares estimation can be used to extract valuable characteristic data from multiple information data. It is supposed the smooth and random information signal containing noise is:

$$F(x)=\alpha(x)+\beta(x) \tag{1}$$

Among them, $\alpha(x)$ represents target information data, and $\beta(x)$ represents noise information data. Denoising the image information is to minimize the difference between $\alpha(x)$ and $\delta(x)$, from which the Formula (2) can be obtained:

$$M(x)=\alpha(x)-\delta(x) \tag{2}$$

$M(x)$ minimum represents the minimum expected value of noise removal.

Unlike Wiener filtering, median filtering has strong applicability. Wiener filtering requires a stable data signal as a prerequisite, while median filtering is not affected by the signal. Its working expression is:

$$E(x,y)=med\{e(x-m,y-n),(m,n\in S)\} \tag{3}$$

Among them, $e(x,y)$ represents the image information before denoising, and $E(x,y)$ represents the image information after denoising.

Based on the extraction and classification function of artificial intelligence multi image feature fusion model, the performance of the model is required. Therefore, the application of VC dimension theory and structural risk minimization principle is proposed.

① VC dimension theory

The VC dimension reflects the classification and extraction ability of the fusion model. The larger the VC dimension of the fusion model is, the more complex the fusion model is, and the stronger its classification and extraction ability is. It can be used to measure the size of the fusion model, so as to judge the complexity of information data. If there is an information data set U, the VC dimension of set U refers to the number of information data in set U that can correctly classify set N, that is, if there is Formula (4):

$$U_n = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\} \tag{4}$$

Then, U$_n$ in Formula (4) is the largest set that U correctly classifies set N, and any x_k, y_k, $k = 1,2,\ldots,n$ can be assigned a value of 1/- 1 to form a set:

$$N_n = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\} \tag{5}$$

Moreover, there must be a function in set U that can accurately classify N_n, that is:

$$d(x_k) = y_k, k = 1,2,\ldots,n \tag{6}$$

② Structural risk minimization principle

Structural risk minimization is to minimize the data risk and reduce the VC dimension, so that the expected risk of the fusion model on the entire data set can be controlled. The formula is:

$$P(d) \le P_{\min}(d) + \sqrt{\frac{8}{e}(f(\ln\frac{2e}{f} + 1)\ln\frac{4}{\omega})} \tag{7}$$

Among them, d represents the function that can accurately classify N_n in VC dimension. e is the number of elements of the set, and f is the VC dimension of the sample set U, so the trust degree of the structural risk minimization principle on the VC dimension can be expressed as:

$$\Psi\ (f, e, \omega) = \sqrt{\frac{8}{e}(f(\ln\frac{2e}{f} + 1)\ln\frac{4}{\omega})} \tag{8}$$

3. Experiment and Evaluation on Classification of Vocal Music Theme

3.1 Exploration on Intelligent Multi Image Feature Fusion in Music Type

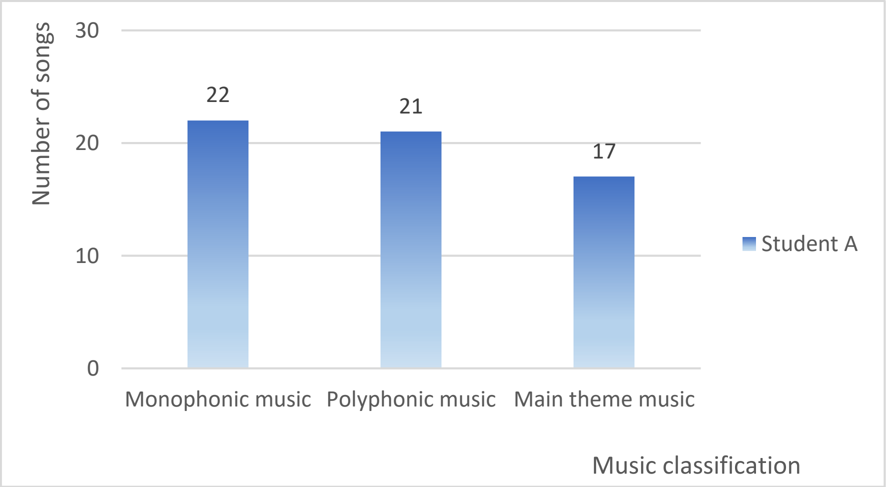

Figure 4: Diagram of the judgment result of student A’s music status

In order to verify the effect of the proposed intelligent multi image feature fusion on the classification of vocal theme, this paper used the artificial intelligence multi image feature fusion model for experiments. 60 pieces of music were selected as the research data, and then 3 students from a music college were selected to judge the music status of the three pieces of music. Figure 4, Figure 5 and Figure 6 are the judgment results of the three students.

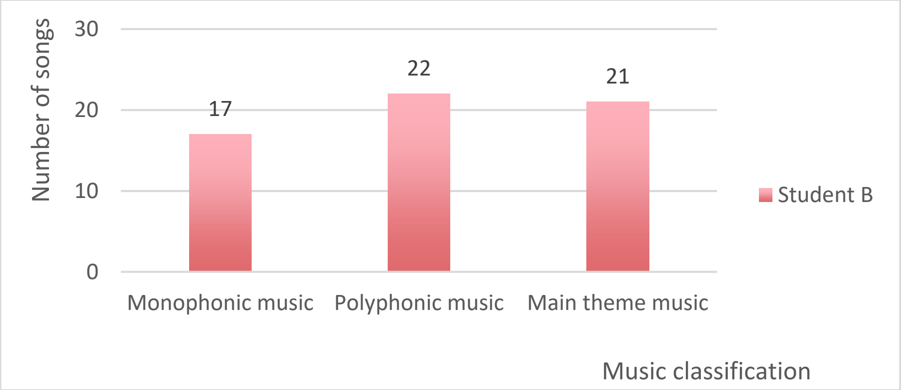

Figure 5: Diagram of the judgment result of student B’s music status

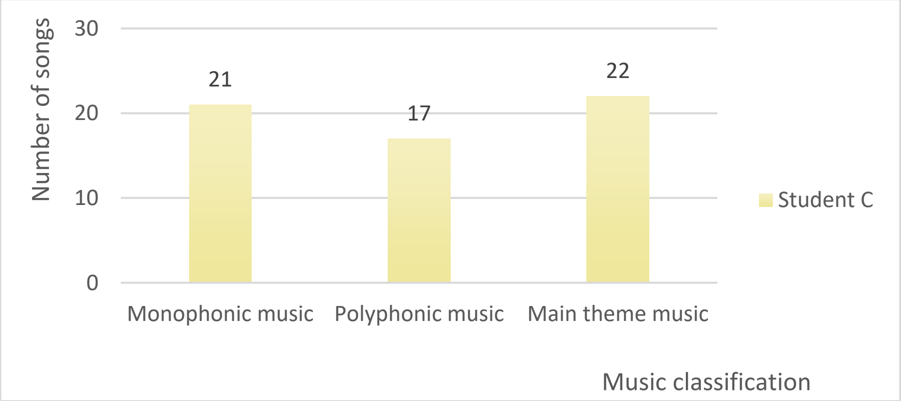

Figure 6: Diagram of the judgment result of student C’s music status

It can be seen from Figure 4 that Student A’s judgment result was that he thought the number of monophonic music was 22, the number of polyphonic music was 21, and the number of theme music was 17. Figure 5 shows the judgment result of student B. In his judgment, the number of monophonic music was 17; the number of polyphonic music was 22, and the number of theme music was 21. The judgment result of student C can be seen from Figure 6. Student C thought that the number of monophonic music was 21; the number of polyphonic music was 17, and the number of theme music was 22. The judgments of the three students were different, so this paper used the artificial intelligence multi image feature fusion model to judge the 60 pieces of music with its multiple sensors. The data results are shown in Figure 7.

The data in Figure 7 shows that the number of monophonic music was 17; the number of polyphonic music was 22, and the number of theme music was 21. The data of this result was consistent with the judgment data of student B. Therefore, it was preliminarily believed that student B’s judgment was correct, and the number of music was classified according to the results of student B.

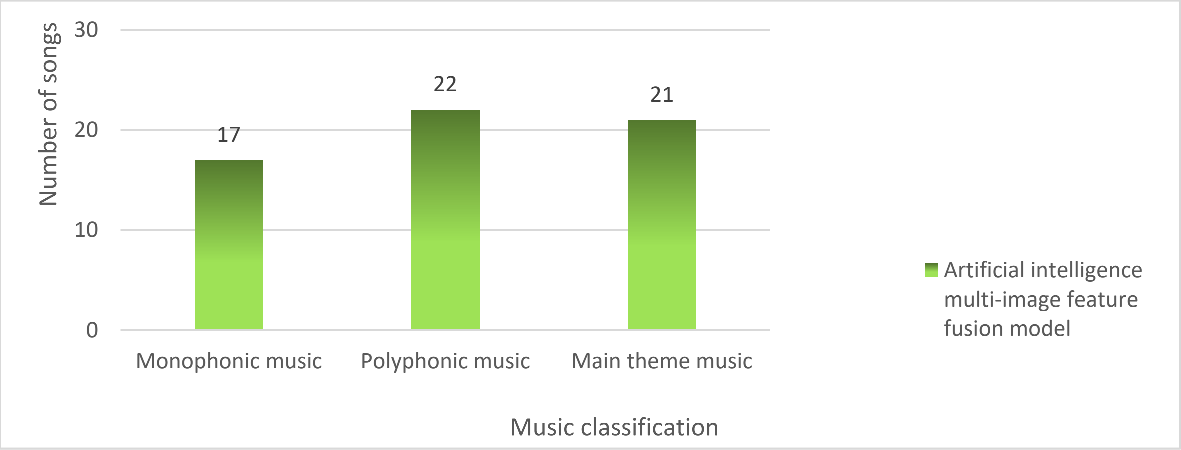

Figure 7: Judgment result of artificial intelligence multi-image feature fusion model

In order to determine the accuracy of the results, this paper consulted two professional music teachers in this conservatory, and compared the results of music teachers' judgment with the classification results of the intelligent multi image feature fusion model. The results are shown in Figure 8.

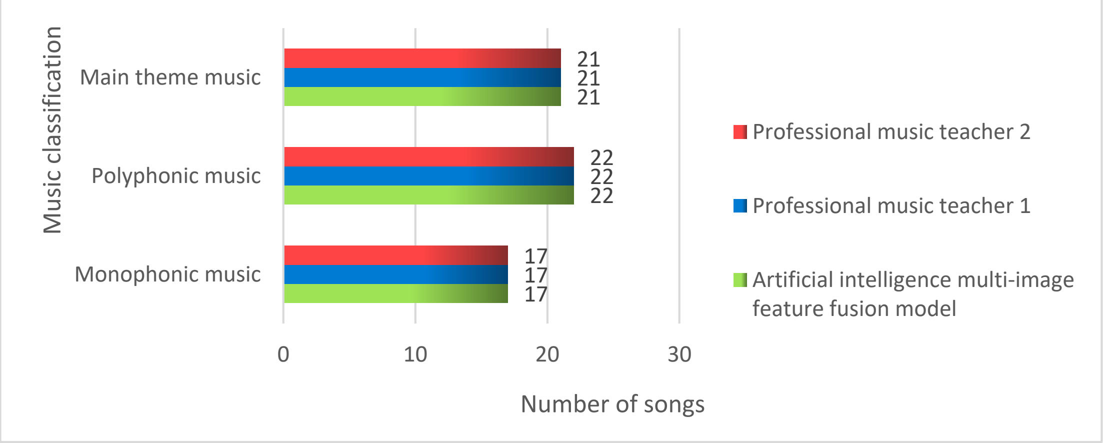

Figure 8: Comparison chart of classification results

The experimental results showed that the number of monophonic music was 17; the number of polyphonic music was 22, and the number of theme music was 21. It was consistent with the judgment result of AI multi image feature fusion model, that is, consistent with student B's judgment, which fully demonstrated the accuracy of AI multi image feature fusion model for music classification.

3.2 Evaluation on the Role of AI Multi Image Feature Fusion Model

In order to more carefully verify the positive role of the intelligent multi-image feature fusion model in promoting the classification of the main melody of vocal music, this paper mixed and cut the main melodies of these 60 pieces of music. These 60 pieces of music cover six types of vocal classification (tenor, baritone, bass, soprano, mezzo-soprano, contralto). The music clips are mixed to obtain a total of 140 clips, male voice clips and female voice clips each account for half, and then the intelligent multi-image feature fusion model is used to determine the vocal classification of this music, and the results are shown in Figure 9.

From Figure 9, the vocal type coverage of the mixed-cut music can be intuitively and vividly understand. Figure 9A shows the melody segments for tenor, baritone and bass, and Figure 9B shows the melody segments for soprano, mezzo-soprano and alto. From the comparison, it can be found that in this mixed-cut music, the bass has the most content, the data line shown in the picture is the longest, the alto has the least content, and the data line is the shortest. Data fluctuations in the graph represent changes in vocal genre. It can be seen that the artificial intelligence multi-image feature fusion model can not only display the vocal type of the mixed-cut music more vividly and intuitively, but also improve the classification efficiency, making the classification of the main melody of vocal music simple and convenient.

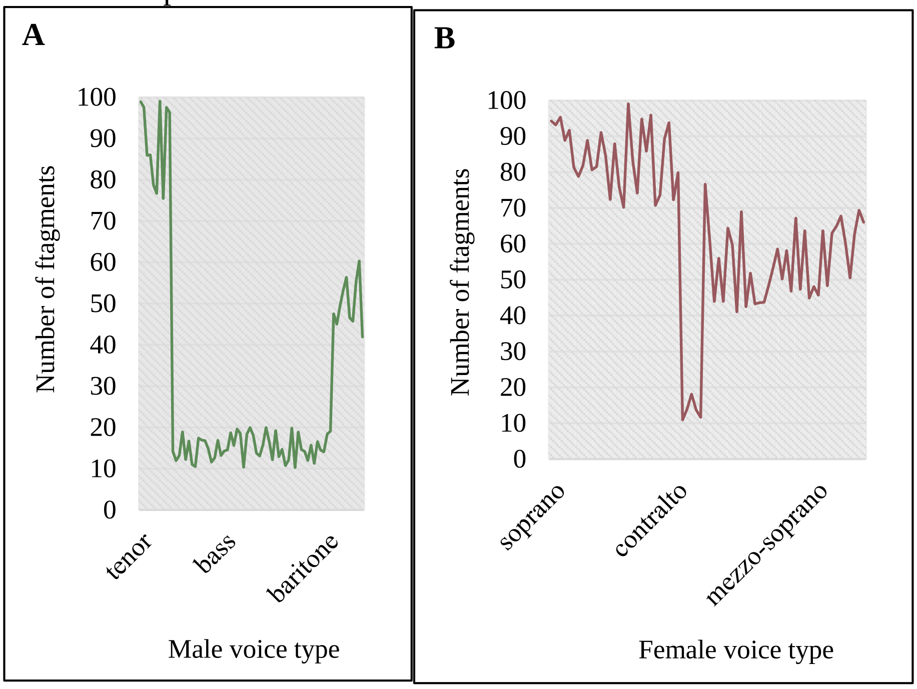

Figure 9A is the melody fluctuation diagram of the male voice Figure 9B is the melody fluctuation diagram of the female voice

Figure 9: Vocal main theme classification judgment result

Multi image feature fusion model based on artificial intelligence can use multiple sensors to sort out the main melody clips, and then classify them in the form of pictures. In order to verify the help of the artificial intelligence multi image feature fusion model to promote the classification of vocal theme in music learning, this paper selected 4 volunteers who do not have a sound to do the experiment. Before the experiment, four volunteers were given a scientific knowledge of the six types of vocal music, and then the four volunteers were asked to listen to the mixed clip music to judge whether the mixed clip music covered six types of vocal music. The result was not satisfactory. After that, the images obtained from the artificial intelligence multi image feature fusion model were shown to four people, and mixed cut music was played while watching, so that it can judge the classification type of vocal music theme again. The accuracy of the results obtained has been greatly improved. Figure 10 shows the comparison of the accuracy before and after the test.

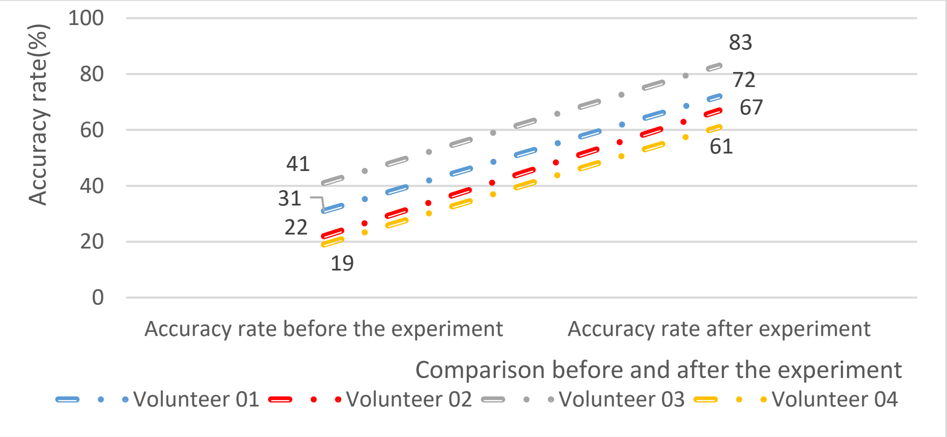

Figure 10: Comparison chart of the accuracy rate before and after the test

It can be seen from the chart data that the correct rate of the four volunteers' judgment had risen in a straight line. It can be seen that the AI multi image feature fusion model plays a positive role in the classification of vocal music themes. It can not only improve the convenience of the classification of vocal music themes, but also enhance people's understanding of vocal music, thus improving their ability to identify music, and improving their taste and aesthetics.

4. Conclusions

The importance of music on the road of human development can be seen. Intelligent multi image feature fusion is of positive significance for improving the classification of vocal music themes. It can be seen from the experimental data that intelligent multi image feature fusion plays an extraordinary role in promoting the classification of vocal music themes. It is also an inevitable trend to introduce it into it, which is also crucial for the analysis of music, the sublimation of emotion and aesthetics. However, due to the time constraints, the experimental results may have some errors due to the short duration of this experiment. Moreover, the technical requirements of intelligent multi image feature fusion are high, and professional talents need to be trained to apply it to music. There is a large demand for social talents. However, with the sustainable development of society, the solution of these problems would be in sight. In order to further understand the impact of intelligent multi image feature fusion on the classification of vocal theme, the analysis results would be further improved in the subsequent research.

References

[1] Lee, Deborah, Lyn Robinson, and David Bawden. "Modeling the relationship between scientific and bibliographic classification for music." Journal of the Association for Information Science and Technology 70.3 (2019): 230-241.
[2] Demorest, Steven M., Jamey Kelley, and Peter Q. Pfordresher. "Singing ability, musical self-concept, and future music participation." Journal of Research in Music Education 64.4 (2017): 405-420.
[3] Birajdar, Gajanan K., and Mukesh D. Patil. "Speech/music classification using visual and spectral chromagram features." Journal of Ambient Intelligence and Humanized Computing 11.1 (2020): 329-347.
[4] Paquette, Sebastien. "Cross-classification of musical and vocal emotions in the auditory cortex." Annals of the New York Academy of Sciences 1423.1 (2018): 329-337.
[5] Hu, Xiao, Kahyun Choi, and J. Stephen Downie. "A framework for evaluating multimodal music mood classification." Journal of the Association for Information Science and Technology 68.2 (2017): 273-285.
[6] Rosner, Aldona, and Bozena Kostek. "Automatic music genre classification based on musical instrument track separation." Journal of Intelligent Information Systems 50.2 (2018): 363-384.
[7] Nam, Juhan. "Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach." IEEE signal processing magazine 36.1 (2018): 41-51.
[8] Simones, Lilian Lima. "A framework for studying teachers' hand gestures in instrumental and vocal music contexts." Musicae Scientiae 23.2 (2019): 231-249.