Scholar Publishing Group
International Journal of Art Innovation and Development
https://doi.org/10.38007/IJAID.2023.040201
ISSN 2789-7192 Vol. 4, Issue 2: 1-8

Optimizing Ethnic Vocal Performance Using Artificial Intelligence Technology

Yingyan Zhang[1,2]

[1]*Xinghai Conservatory of Music, Guangzhou, China*

[2]*Center for International Education, Philippine Christian University, Manila, Philippines*

ziyubaby1021@163.com

Keywords: Ethnic Vocal Performance, Artificial Intelligence Technology, BP Neural Network, Acoustic Indicators

Abstract: The regional characteristics of Chinese folk songs are a unique artistic expression with a long history of Chinese ethnic culture. How to optimize the performance of ethnic vocal music is a concern for people. The Conservatory of Music admits 36 ethnic vocal performers aged 18-22 who specialize in music to perform. Senior professional music teachers subjectively evaluate the quality of songs on a 10 point scale based on their rich vocal knowledge and experience. Based on the characteristics of ethnic vocal performance, this article evaluated the singing effect in ethnic vocal performance using BP (Back Propagation) neural network using three selected acoustic indicators: average energy, frequency error, and range error. The average subjective evaluation score in this article was 8.4, and the average neural network evaluation score was 8.9. This article would provide a basis for the objective evaluation of singing sound quality in ethnic vocal performance using artificial intelligence technology.

1. Introduction

The combination of singing techniques and singing expression is another important principle in vocal performance art. The technique of singing and the expressive power of music are two indispensable factors in vocal performance art. Without technique, there is no expressive ability; on the contrary, if there is a lack of artistic expression, singing techniques would be meaningless. Therefore, the perfect combination of singing techniques and artistic expression is another important principle in vocal performance art. Singing techniques play a crucial role in vocal performance, but simply using them is not enough. Only by perfectly combining techniques with artistic expression can their effectiveness be fully realized. The principle of unity between singing

techniques and artistic expression is applicable in every era, as well as to various styles and forms of singing works. Only by mastering this principle well can the artistry of vocal performance be better reflected. Under the background of artificial intelligence technology, the singing art of ethnic vocal music presents stable characteristics such as nationality, scientificity, artistry, and modernity. This relatively stable characteristic is very valuable for cultivating talents in ethnic vocal art, grasping the development direction of ethnic vocal art, and handling the relationship between inheritance and innovation.

At present, there has been a deep accumulation of research on the theory and practice of traditional ethnic vocal art in China, including singing techniques and forms of expression, as well as attention to the spirit of the times and aesthetic taste. Danylets V aimed to describe the structure and stylistic components of performing folklore, as well as reflections on existing definitions of various concepts of performing folklore [1]. Zhou Y analyzed the vocal culture of modern China. It was formed by the interaction between national and international cultural models, which was mainly formed by language and music language [2]. Wahyono Wahyono aimed to analyze the strategy of performance art, especially as the source and identity movement of Minangkabau traditional dance movement, and as the marketing media of tourism industry in West Sumatra Province in particular [3]. Jude Gretchen suggested understanding this performance practice as participating in plasma sound, which was a purely additive phenomenon different from sound+electricity [4]. Heniwaty Yusnizar believed that the economy was steadily growing, and technology was driving the rapid development of various industries. Art education was increasingly receiving social attention, and vocal teaching was also receiving attention and love from parents and students [5]. Their research does not fully reflect artistry.

This article analyzed the principles, characteristics, artistic means, and artistic expression of vocal performance from the perspective of vocal performance aesthetics. It explored the methods and approached to enhance the artistry of vocal singing through studying vocal art aesthetics, providing reference for the learning and reference of vocal singers.

2. Artificial Intelligence Technology Optimizes Ethnic Vocal Performance

2.1. Artificial Intelligence

Artificial intelligence is a comprehensive science that integrates psychology, philosophy, and computer technology. Although there is also “intelligence” in it, it is not an entity. It is an artificial intelligence device that can simulate all human behavior through intelligent technology. It’s like giving machinery a person’s intelligence to do many things [6-7].

At present, computer music is developing at an astonishing speed, accompanied by the emergence of artificial intelligence. Music encompasses two important aspects: One is composition, and the other is performance. Different genres of music convey different styles, and performers need to inject different rhythms and dynamics into it to produce rich expressive power. The development of singing style transformation technology has provided new ideas for music style transformation. Its research can provide new ideas for music creation and new ideas for music creation, which is an important content in many artificial intelligence fields [8].

Music information retrieval is an emerging technology generated in the massive digital music environment, which combines computers with music and has broad application prospects [9-10]. In recent years, due to breakthroughs in artificial neural networks, deep production models have become one of the main technologies for automatic music generation. This model can generate new data, modify existing data in principle, and achieve attribute transfer of existing data.

Artificial intelligence, as an emerging industry, plays an important role in economic and social development. Artificial intelligence with the ability to “create” is an emerging industry in the field

of artificial intelligence, and its comprehensive development would promote artificial intelligence to "create" more songs. With the increasing application of artificial intelligence in human life, people suddenly realize that, unconsciously, artificial intelligence has integrated into literature and art, and can even "create" music, painting, literature, and so on that were originally believed to be only achievable by humans, and the scope of this integration continues to expand [11].

After extensive learning and research on thousands of excellent music works in the music industry, they can be handed over to artificial intelligence for "self-learning", thus enabling them to have a certain understanding of basic elements such as melody characteristics, music style, and composition methods of music, and then "create" according to a certain program.

In ethnic vocal performance, signal pre emphasis is achieved through a digital filter that enhances high-frequency characteristics. The transfer function of this filter is as follows:

$$H(z) = 1 - \varepsilon M^{-1} \tag{1}$$

Among them, ε is the pre emphasis coefficient. It is assumed that the time-domain waveform of the voice signal is $x(n)$:

$$x(n) = \sum x(m)x(m+k) \tag{2}$$

Only by comprehensively understanding the work from both emotional and rational perspectives can the theme of the work be correctly conveyed to the audience in vocal performance. All music art is formed in a certain context of life and culture. To truly grasp the music in this context, it is necessary to understand and feel its life and culture as much as possible, so as to reflect the spiritual meaning of the work. The genre characteristics of Chinese folk songs are closely related to the local regional characteristics. Therefore, the style of music is correspondingly graceful and elegant. In vocal performance, it is necessary to integrate it with cultural backgrounds and phenomena, and to intersect and integrate it with other disciplines, so as to achieve an accurate grasp of its thematic ideas [12].

The channel filter $S(Z)$ and prediction error filter $y(Z)$ are as follows:

$$S(Z) = G/(1 - \sum aZ) \tag{3}$$

$$y(Z) = 1 - \sum aZ \tag{4}$$

2.2. Ethnic Vocal Performance

As one of the oldest cultural forms, vocal performance is a manifestation of people's cultural mentality and spiritual pursuit of social life. Vocal performance art is a musical expression method based on auditory aesthetics and assisted by visual aesthetics, belonging to the category of music performance art. The content of vocal performance art involves multiple aspects such as music, literature, language, history, drama, etc. It is a comprehensive art. Vocal performance art is an activity of secondary creation based on the original work and combined with one's own understanding and interpretation. This requires adherence to the principles of unity between objective authenticity and subjective initiative, unity between historical style and contemporary taste, and unity between singing techniques and artistic expression. Vocal performance art is a manifestation of a singer's comprehensive qualities and abilities. It requires singers not only to have excellent singing skills, but also to have high artistic literacy, elegant typhoon style, good performance psychology, and rich practical experience [13]. This article would extract the acoustic parameters of artistic voice signals and objectively evaluate the process using neural networks, as shown in Figure 1.

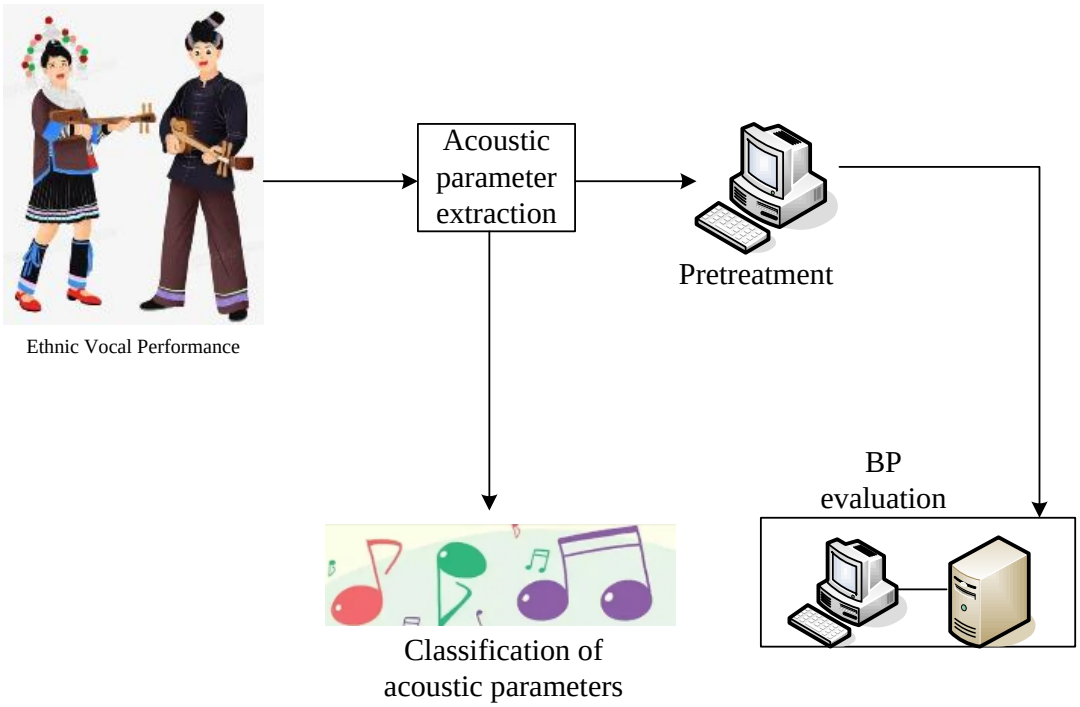

Figure 1. Acoustic parameter extraction and objective evaluation process of neural network for artistic voice signals in this article

There are various types of ethnic vocal works in China, and performers must choose appropriate singing methods based on their own actual situation. The absorption of ethnic singing art from Chinese theatrical performances has significant practical significance for the training of "emotion", "spirit", and "form" in current ethnic vocal singing. However, there is one issue that must be taken seriously, which is the unity of physical performance and singing tones. Vocal art requires the comprehensive development of singing and performance. However, at any time, singing should occupy the main position, and the purpose of physical performance is to make singing more lively. It is only an auxiliary means, with a distinction between primary and secondary importance, and cannot be dominated by others. Otherwise, it is not a singing art. Therefore, it is necessary for everyone to have a good control over the scale and quantity of performances, neither focusing on physical form nor on physical form, in order to achieve a harmonious unity between singing and performance [14-15].

Vocal singing is an art form that can directly express human emotions and stimulate people's imagination. It has developed to this day and has spawned many different categories. Vocal performance is a very important form of musical art expression. From the perspective of vocal performers, if one wants to present a perfect live performance, it is necessary to analyze the theoretical level of the work, integrate multiple factors, and conduct in-depth thinking [16]. Motivation is the cause of behavioral arousal, so it is necessary to conduct in-depth analysis and interpretation of the performed work. Based on the interpretation of the work, the first step is to create a performance context that can promote one's own desire to perform and drive one's own performance. Only then can it have an impact on the audience [17]. Various types of vocal works have their own unique style characteristics, and the formation of these characteristics is influenced by factors such as national history and culture, ethnic customs, music genres, and the personal style characteristics of creators and singers. In order to accurately create a performance context, vocal singers must have a clear understanding of the style of the work and then explore the expressive

characteristics of the work. Only in this way can the unique charm of the work be better demonstrated. Before singing a song, one should first be able to read the lyrics correctly and ensure that the words are accurate and round. During the process of reading the lyrics, one can experience the tone and charm, thereby gaining a deeper understanding of the meaning of the words [18].

3. Exploration Experiment on Optimizing Ethnic Vocal Performance

3.1. Ethnic Vocal Performance Experiment

The College of Music of the University recruits 36 ethnic vocal performers aged 18-22 who are majoring in music. The educational requirements are: at least 2 years old, without throat diseases or upper respiratory diseases. The sound sampling location should be located in a soundproof room, and the surrounding noise should not exceed 45 decibels. Recording is done using a microphone connected to a computer sound card, with a computer sampling frequency of 48 KHZ. The recorded signal is stored on a sound file storage card. Recording is conducted at a distance of 10 centimeters from the microphone, thus recording professionally trained ethnic singing information. At the same time, senior professional teachers from the Conservatory of Music would subjectively evaluate the quality of songs on a 10 point scale based on their rich vocal knowledge and experience.

3.2. BP Neural Network Evaluation

The average energy of singing, frequency error, and range error were selected as evaluation parameters. The average energy of singing, frequency error, and range error are normalized, and then used as input parameters. The scores of senior professional teachers' subjective evaluation of singing quality are used as output parameters to establish a non-linear relationship between singing quality scores and evaluation parameters in ethnic vocal performance.

4. Optimization Results of Ethnic Vocal Performance

On this basis, through the analysis of acoustic characteristics. The acoustic characteristics were analyzed, and the acoustic characteristic values of the acoustic characteristics were obtained, and the acoustic characteristic values were analyzed. The basic frequency is different between men and women. The basic frequency for women is approximately 366 Hz to 496 Hz, while the basic frequency for men is approximately 138 Hz to 258 Hz. From the scatter plots of the first and third resonance peaks of the signal, it could be seen that the first resonance peak of the singing signal was around 317-596 Hz, while the third resonance peak was around 1741Hz-2339 Hz. The fundamental frequency and vibration peak of the singing art voice signal are shown in Figure 2.

Every song has its own thoughts and emotions to express, reflecting a pursuit of the times and a pursuit of life. Therefore, a vocal singer must first grasp the theme of a song when obtaining it. In terms of artistic expression, vocal music presents a non visual and non semantic characteristic. Therefore, vocal performers can improve their artistic imagination by learning and learning other literary works, thereby making their vocal performance more meaningful and artistic.

BP neural network belongs to the "backward" network, and the number of hidden layer neurons directly affects the performance of the network. In this article, the number of neurons should be between 3 and 10. Starting from the minimum number of three neurons, the number of cells increased one by one. It was validated to the maximum number of 10 neurons, thus determining that the number of neurons corresponding to the minimum training error was the optimal number of hidden layer neurons. The training error of the number of neurons in different hidden layers is shown in Table 1.

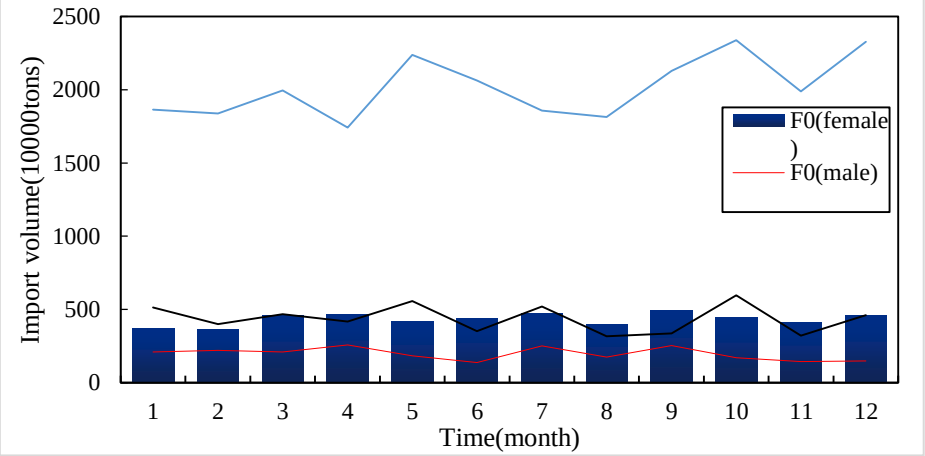

Figure 2. Basic frequency and vibration peak of singing art voice signal

Table 1. Training error of the number of neurons in different hidden layers

Number of hidden layer neurons	Training error
3	0.0512
4	0.0412
5	0.0432
6	0.0433
7	0.0356
8	0.0223
9	0.0116
10	0.0516

The “sound”, “emotion”, “spirit”, and “form” of singers are an organic whole, and traditional Chinese drama can better grasp this organic unity. Therefore, learning and borrowing these organic and unified singing methods is of great significance for promoting the development of Chinese vocal art. It is hoped to attract the attention of vocal learners and provide corresponding learning in opera performance. If possible, schools can offer relevant courses, which would be of great help in folk song singing and performance. The regionality of Chinese folk songs is a long-standing form of artistic expression in Chinese history and culture. Regionality is a major characteristic of folk song creation, and it is also an important factor in the popularity of Chinese folk songs among the general public. Therefore, in grasping the local ethnic style, it is a practical and effective method to represent different types of regions and regions. A detailed analysis, introduction, and summary of various color regions were conducted from the perspectives of geographical environment, modes, sound sequences, and rotation techniques, so as to facilitate vocal learners to better grasp them. When learning ethnic vocal music, it is necessary to accurately grasp the characteristics of Chinese language and music, absorb nutrients from various aspects, and improve one’s musical literacy, artistic cultivation, and cultural accumulation. To master the ethnic style in folk song singing, it is necessary to maintain a close relationship with traditional ethnic music for a long time.

This article selected three acoustic indicators: average energy, frequency error, and range error, and evaluated the singing effect in ethnic vocal performance using BP neural network. 10 songs were randomly selected as the training set, and the remaining 8 songs were used as test samples. The objective evaluation score is shown in Figure 3, and the deviation from the subjective evaluation score of senior professional teachers was within 4%. Therefore, a good evaluation of singing quality could be achieved. The average subjective evaluation score was 8.4, and the average

neural network evaluation score was 8.9.

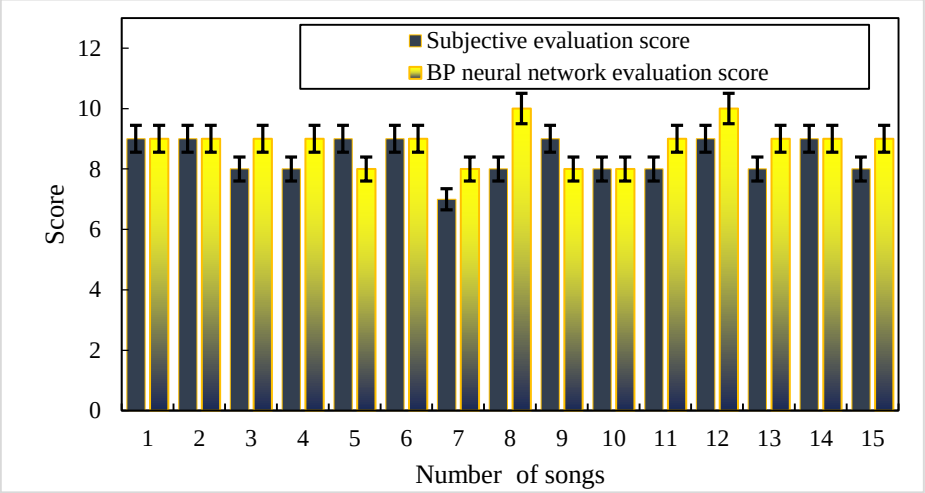

Figure 3. Objective evaluation score

5. Conclusions

It is a practical and effective method to grasp the local ethnic style, based on the regional characteristics of different types and regions, and to have a representative grasp of it. A detailed analysis, introduction, and summary of various color regions were conducted from the perspectives of geographical environment, sound arrangement, and rotation techniques, so as to facilitate vocal learners to better grasp. This article used BP neural network to objectively evaluate it and conducted a preliminary analysis of the singing effect in ethnic vocal performance. It is helpful to improve the quality and training methods of vocal talents; This plays a certain role in understanding the changes in sound indicators and objectively evaluating artistic sound. There is huge room for improvement in the establishment of a singing sound quality database and the selection of features for singing sound quality evaluation in the future.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Danylets, V. . "The hutsul music features in the structural and stylistic context of the performing folkloryzm." Problems of Interaction Between Arts Pedagogy and the Theory and Practice of Education 57.57(2020):77-88.

[2] Zhou, Y. . "Verbal aspects of China's vocal art system." Aspects of Historical Musicology 21.21(2020):137-149.

[3] Wahyono, Wahyono, and Benny Hutahayan. "Performance art strategy for tourism segmentation:(A silat movement of Minangkabau ethnic group) in the event of tourism performance improvement." Journal of Islamic Marketing 11.3 (2020): 643-659.

[4] Jude, Gretchen. "Vocal Performance Through Electrical Flows: Making Current Kin." Performance Philosophy 4.2 (2019): 393-409.

[5] Heniwaty, Yusnizar. "Saman Dance of The Aceh People: Identity and Actualization." Journal of Community Research and Service 2.1 (2018): 184-191.

[6] Fu, Lei. "Discussion on the differences between theory and practice in vocal music teaching." Region-Educational Research and Reviews 3.1 (2021): 6-9.

[7] Swan, Elaine, and Rick Flowers. "Lasting impressions: Ethnic food tour guides and body work in southwestern Sydney." Gender, Work & Organization 25.1 (2018): 24-41.

[8] Nugroho, Kristiawan, and Edi Noersasongko. "Enhanced Indonesian ethnic speaker recognition using data augmentation deep neural network." Journal of King Saud University-Computer and Information Sciences 34.7 (2022): 4375-4384.

[9] Bernabo, Laurena. "Whitewashing diverse voices:(De) constructing race and ethnicity in Spanish-language television dubbing." Media, Culture & Society 43.7 (2021): 1297-1310.

[10] Konovalova, Iryna Yu. "The composer phenomenon in a historical retrospect of an authorship idea formation in European musical culture of the middle ages." Linguistics and Culture Review 5.S2 (2021): 284-302.

[11] Erwenta, Jona, Leo Agung, and Sunardi Sunardi. "The Values of Character Education in the Didong Art Performance: A Study of Enculturation Process in Gayonese Society." International Journal of Multicultural and Multireligious Understanding 5.4 (2018): 196-203.

[12] Amer, Amena. "Between recognition and mis/nonrecognition: Strategies of negotiating and performing identities among white Muslims in the United Kingdom." Political Psychology 41.3 (2020): 533-548.

[13] Anderson, Amy, and Darren Byler. "“eating hanness”: Uyghur Musical tradition in a time of re-education." China Perspectives 2019.2019-3 (2019): 17-26.

[14] Van Elferen, Isabella. "Dark timbre: the aesthetics of tone colour in goth music." Popular Music 37.1 (2018): 22-39.

[15] Shafazhinskaya, Natalya Evgenyevna, et al. "Integration of traditions in crossover music as a way to bridge the intergenerational gap." Humanities & Social Sciences Reviews 7.6 (2019): 1230-1234.

[16] Oram, Daron. "De-colonizing listening: Toward an equitable approach to speech training for the actor." Voice and Speech Review 13.3 (2019): 279-297.

[17] Zixuan, Li. "Research on the current situation of Chinese national vocal music education from the perspective of educational psychology." education 3.6 (2021): 48-58.

[18] Rangarathnam, Balaji, Towino Paramby, and Gary H. McCullough. "“Prologues to a bad voice”: effect of vocal hygiene knowledge and training on voice quality following stage performance." Journal of Voice 32.3 (2018): 300-306.