

Research on the Extraction of Digital Features of Vocal Performance Styles and the Construction of a Database for Inheritance and Protection

Qiong Duan¹

¹Conservatory of music, Shanxi College of Applied Science and Technology, Taiyuan 030062, Shanxi, China

Corresponding author: Qiong Duan (e-mail: a18536613033@163.com)

ABSTRACT With the paradigm shift in vocal art research, transforming highly abstract and experience-dependent vocal performance styles into quantifiable and analyzable digital objects has become an important issue in music information processing and digital heritage preservation. This study focuses on digital feature extraction and database construction for vocal performance styles, aiming to build a multi-dimensional feature fusion framework and a searchable database system. The study first establishes a multimodal data acquisition standard covering audio signals, physiological parameters, and body posture movements. Then, short-time Fourier transform, fundamental frequency tracking, surface electromyography analysis, and computer vision technology are used to deconstruct and quantify performance styles from acoustic, physiological, and physical dimensions. Based on these features, a vocal performance style database supporting multidimensional retrieval, similarity matching, and visual analysis is designed and implemented. Experimental results show that vibrato frequency, formant position, harmonic-noise ratio, energy dynamic range, and posture-related features can effectively distinguish different vocal schools and individual performance characteristics. The database achieves an average response time of 1.32 seconds under a concurrent retrieval environment, indicating acceptable retrieval efficiency for teaching and research applications. The proposed framework provides an operable technical path for the digital preservation of vocal art and offers data support for vocal teaching, style comparison, and AI-assisted creation. Its time-frequency feature extraction and waveform-indexing mechanism also provide methodological reference for other wave-based signal archives, including acoustic and electromagnetic signal data management.

INDEX TERMS vocal performance style, digital feature extraction, multimodal database, time-frequency analysis, waveform signal processing

I. INTRODUCTION

VOCAL art, as an important part of human intangible cultural heritage, embodies rich historical information and artistic wisdom within its performance styles, mirroring the intricate material morphology and complex material spatial distributions that characterize the evolution of the industry. The systematic preservation of these vocal nuances parallels the digital modeling of antique looms and industrial textile heritage, ensuring that both abstract performance and physical manufacturing excellence are captured as quantifiable cultural assets. However, the forms of vocal performance have long leaned on the oral tradition and subjective aesthetical judgement with experiences. This is a highly abstract and difficult to quantify characteristic that equips it with the dual risks of transmission disruption and memory loss in the digital age.

Converting the vocal performance styles of various schools and historical periods into searchable and analyzable digital objects—much like bridging the semantic gap between the geometric physical data of antique and the complex material machines morphology of industrial heritage—has become a major challenge in the interdisciplinary field of musicology and information science. This digital transformation requires the same DT-KG deep fusion methodologies used to quantify material spatial distribution and manufacturing excellence to ensure the precise preservation of abstract cultural assets.

In recent years, the fast development of artificial intelligence and digital humanities has brought us technological possibilities of leaping of vocal art from “audible” to “visual and calculable.” The performance style of vocal works is not only manifested in basic musical elements such as pitch,

rhythm and timbre, and permeates the physiological characteristics of the singer's voice, the physical movement and the emotional expression, which has a complex and multi-dimensional structure. Traditional audio analysis methods are often restricted to the extraction of single-dimensional acoustic features, which will make it difficult to comprehensively reflect the overall characteristics of performance style; while relying on subjective descriptions of the data combined with manual annotation is not enough to satisfy the needs of a large-scale and standard data construction. Therefore, determining a digital extraction framework that can integrate multi-dimensional features and combine the subjectivity of quantification with the subjectivity of art, and constructing a database of vocal performances style, which is based on the framework, has great theoretical and practical value. This framework bridges the semantic gap by correlating low-level physical descriptors (F0, EMG) with high-level artistic annotations—such as ‘vocal tension’ or ‘cultural timbre’—through a hybrid expert-scoring and data-mapping system, thereby translating raw numbers into quantifiable artistic wisdom.

This study focuses on the performance style of vocal works, and seeks to answer three main questions: First, how to construct a multi-dimensional feature extraction index system for vocal performance style to obtain style-related features from audio signals, physiological signals, and video images simultaneously; second, how to construct a database architecture to support storage, retrieval, and visualization analysis of

multi-dimensional features; third, how effective the constructed database is in practical application scenarios such as vocal performance style recognition, heritage preservation, and teaching assistance, etc. To answer these questions, in this paper, the methodology, experimental verification and system implementation, are discussed step by step.

II. RELATED WORK

Research on vocal performance art presents a paradigm shift from humanistic interpretation to scientific empiricism and from aesthetic speculation to data-driven approaches. Pechenyuk et al. [1] mapped the characteristics of post-modern art onto contemporary vocal performance practice cases, conducted in-depth textual analysis and phenomenological interpretation, and explored the evolution of vocal works in terms of creative concepts, performance styles and aesthetic reception. Drozhzhina et al. [2] explored how popular vocal music, as an art form with great dissemination power and mass base, influences and promotes the cultural integration and social cohesion of future society. Moura et al. [3] explored the differences in the perception of the range of physical movements of performers by listeners or judges under different music styles, and its relationship with performance evaluation. McNeil et al. [4] confirmed the positive correlation between perfectionism and musical performance anxiety. Nara [5] focused on understanding how K-pop appropriates and reshapes the traditional Korean

male image to adapt to the aesthetics of the global market. Brown et al. [6] explored whether there is a common optimization principle in pitch organization in traditional vocal music worldwide. Scherbaum and Mueller [7] revealed the transmission process between singers from precise pitch and rhythm coordination to deeper unconscious physiological synchronization. Pearson and Pouw [8] conducted refined motion capture and acoustic analysis on videos and audio of Carnatic music performances. Sharma et al. [9] explored how to improve the accuracy and robustness of automatic classification of music genres by integrating multiple types of features. Karthik et al. [10] systematically discussed and compared different feature extraction methods and evaluated their application effects in MIR tasks. This paper proposes a study on the digital feature extraction and inheritance protection database of vocal performance styles, aiming to build a digital extraction framework that integrates multi-dimensional features, and on this basis, to build a searchable, analyzable, and inheritable vocal performance style database, promote vocal performance research from empirical description to data-driven, from individual interpretation to systematic inheritance, and provide theoretical and practical support for the digital protection and innovative development of vocal art.

III. METHODS

A. AUDIO DATA ACQUISITION AND PREPROCESSING

The first step in extracting digital features of vocal performance style is to propose a preliminary multimodal data collection framework. While the current study utilizes a group of 45 professional singers, it is recognized that this sample serves as an initial reference rather than a definitive universal standard. This pilot dataset provides a foundational baseline for the five major singing styles, acknowledging that broader stylistic variations will require significantly expanded N-counts in future research phases. This study selected five representative vocal styles as data collection objects, including Western bel canto, folk singing, popular singing, opera singing and original folk singing. A total of 45 professional singers were invited, including 22 males and 23 females, aged between 25 and 65 years old, and all had more than 8 years of singing experience to ensure the maturity and stability of their styles. Each singer sang three representative songs in a standard recording environment, and three types of data, namely audio signals, physiological signals and video images, were collected at the same time [11].

Audio signal acquisition used a Neumann U87 condenser microphone with a sampling rate of 96kHz and a bit depth of 24-bit to ensure the full preservation of high-frequency overtones. While the level of ambient noise was maintained below 20dB(A) and the reverberation time RT60 was kept below 0.3 seconds to ensure signal purity for baseline extraction, supplementary recordings were conducted in acoustically representative environments (e.g., concert halls for bel canto and open-air settings for original folk singing)

to account for the spatial resonance characteristics inherent to each style's authentic performance practice. Physiological signal acquisition was done using a Delsys Trigno surface electromyography system and simultaneously measured the electromyographic activity (Diaphragm, sternocleidomastoid, and Rectus abdominis muscles) during singing at a sampling frequency of 2000Hz. Video image acquisition used two high-speed cameras, simultaneously recording from the front and side at 45-degree angles, at a frame rate of 60fps and a resolution of 1920×1080, for subsequent extraction of body movement features. This study was approved by the Ethics Committee of Shanxi College of Applied Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. A triple filtering mechanism was used in the data preprocessing phase in order to assure that the data is quality. First of all, segments with accompanying audio interference, environmental noise or singing errors were removed, using spectral analysis. The selection criteria were a signal-to-noise ratio greater than 35dB and an unbroken fundamental frequency trajectory. Second, to ensure the extraction of the 'Energy dynamic range' feature remains authentic, preprocessing was performed on dual tracks: a raw, uncompressed track for the analysis of dynamic contrast and spectral intensity, and a normalized track (utilizing peak normalization rather than 3:1 compression) to standardize fundamental frequency tracking and phoneme alignment across different singers. Finally, a fundamental frequency detection algorithm was applied to filter out stable samples of pitch where the fundamental frequency was unstable by more than 30 cents. After preprocessing, a total of 218.6 hours of effective data duration, including 189.2 hours of audio data, 189.2 hours of synchronized physiological signal data, and 218.6 hours of video footage were obtained.

B. CONSTRUCTION OF MULTIDIMENSIONAL FEATURE EXTRACTION SYSTEM

Based on the preprocessed multimodal data, this study constructs a three-dimensional feature extraction system including an acoustic feature layer, a physiological feature layer and a body feature layer. The acoustic feature layer uses the short-time Fourier transform method to extract frequency domain statistical features. The window function is a Hamming window with a window length of 2048 sampling points and a frame shift of 512 sampling points. The spectrum calculation range is 0 to 22050 Hz. On this basis, the following five types of acoustic features are extracted: First, fundamental frequency related features, including average fundamental frequency, fundamental frequency standard deviation, fundamental frequency dynamic range and vibrato frequency and amplitude; Second, energy related features, including short-time energy mean, energy dynamic range and energy decay rate; Third, formant features, including the frequency and bandwidth of the first to fourth formants; Fourth, harmonic noise ratio, used

to quantify the proportion of breathy components during singing; Fifth, spectral centroid and spectral roll-off point, reflecting the brightness of the timbre and the high-frequency energy distribution [12].

The physiological feature layer primarily extracts the singer's vocal physiological effort from surface electromyography (EMG) signals. The raw EMG signals are subjected to a 50Hz notch filter to remove power frequency interference, and then a bandpass filter is used to extract the effective signal in the 20-450Hz frequency band. Based on this, three indicators are calculated for each measured muscle: root mean square amplitude, integrated EMG value, and median frequency. Root mean square amplitude reflects the degree of muscle activation, integrated EMG value measures the total work done by the muscle per unit time, and median frequency indicates the changes in muscle fatigue state and motor unit recruitment patterns. These indicators are statistically analyzed segmentally according to the singing phrases to obtain the physiological change trajectory during the singing of each phrase.

The body posture feature layer extracts body movement features during the singing process based on computer vision technology. The OpenPose pose estimation algorithm is used to detect human key points in the video frame sequence and extract the two-dimensional coordinates of 25 key points including the head, shoulders, elbows, wrists and torso. Based on the key point coordinates, the following three types of movement features are calculated: first, motion volume features, including displacement velocity, acceleration and range of motion of each key point; second, posture features, including head tilt angle, shoulder opening and closing angle and torso forward tilt angle; third, movement coordination features, which quantifies the temporal synchronization between gestures and body movements by calculating the cross-correlation function between upper limb key points. All the above features are output in time series at a frame rate of 30fps, keeping time aligned with the audio features [13,14].

C. DATABASE ARCHITECTURE DESIGN AND IMPLEMENTATION

The vocal performance style database which is designed according to the principle of multidimensional retrieval and research-oriented knowledge base construction, uses the hybrid architecture of MySQL relational database system and MongoDB document-oriented database system. The relational database stores mainly structured metadata that includes singer basic information (name, gender, age, singing style, relationship between teacher and student), song basic information (song title, geographical origin, historical context, lyrics), and also basic information of the recording (time of recording, location of recording, parameters of equipment). The document-oriented database allows the storage of extracted multidimensional feature vectors and time series data of the features. Each recording record is a corresponding JSON document that has complete

data of acoustic, physiological and physical characteristics.

The system employs a dual-core architecture: the 'Repertoire Sub-database' manages work-centric metadata (title, origin, history), while the 'Singing Style Sub-database' manages artist-centric parameters (lineage, teacher-student relationships, and multi-dimensional feature vectors). These cores are dynamically linked via unique IDs to facilitate multi-modal cross-referencing. The two sub-databases are related using singer IDs and repertoire IDs, so that cross-referencing between works and singing style features can be performed. To provide visualized retrieval and analysis of feature data, the database system has achieved a web-based query interface, which supports feature-based similarity query. Users can choose a reference recording and the system would automatically compute the Euclidean distance or the cosine similarity between other recordings in the database against the reference recording on the chosen feature dimension and list the results of the query in the order of similarity. Feature similarity calculation supports single-dimensional/multidimensional weighted combination with weights customizable by the user according to the research need. Furthermore, the system has feature trend analysis capabilities, discovering the statistical analysis of feature mean, standard deviation, and distribution range of the recording samples from specific schools of thought, times or lineages in form of visual charts.

The database system uses a B/S (Browser/Server) architecture with the back-end server being developed with the Spring Boot framework and the front-end developed with the Vue.js framework for the interactive interface. The database connection pool has a maximum connection limit of 100, which finds concurrent access from many users. To ensure efficient retrieval of the feature data with scale, multi-dimensional indexes are constructed for the feature vectors, and the pre-calculated result is cached for commonly used feature combination queries.

IV. RESULTS AND DISCUSSION

A. ANALYSIS OF VOCAL PERFORMANCE STYLE FEATURE EXTRACTION RESULTS

Based on the constructed multidimensional feature extraction system, this study systematically extracted features from 218.6 hours of collected multimodal data, obtaining 1.892 million acoustic feature parameters, 1.419 million physiological feature parameters, and 3.279 million body posture feature parameters. To verify the ability of the extracted features to represent performance styles, performance samples from 10 singers each from Western bel canto, folk singing, and pop singing were selected for feature difference significance testing.

Table 1 compares the mean and standard deviation of the core acoustic features of the three singing styles. The data shows that Western bel canto has the lowest average fundamental frequency (220.5 Hz) and a standard deviation of 8.7 cents, reflecting its pursuit of a wide vocal range and significant pitch variation. Folk singing has the highest

average fundamental frequency (315.7 Hz) and the fastest vibrato frequency (5.8

Hz), reflecting its aesthetic preference for bright high frequencies and dense vibrato. Popular singing has the largest dynamic range (34.2 dB) and the lowest harmonic noise ratio (12.4 dB), confirming its widespread use of breathy vocals and strong dynamic contrast. All the above differences are statistically significant at the $p < 0.01$ level, indicating that the extracted acoustic features can effectively distinguish the performance styles of different styles.

TABLE 1. Comparison of core acoustic features of different singing styles

Feature parameters	Western bel canto singing style (n=10)	Folk singing style (n=10)	Popular singing style (n=10)	F value	p-value
Average fundamental frequency (Hz)	220.5 ± 35.2	315.7 ± 42.8	245.3 ± 38.6	18.32	< 0.001
Fundamental frequency standard deviation (cents)	42.3 ± 8.7	35.6 ± 7.2	38.9 ± 9.1	6.45	0.004
Vibrato frequency (Hz)	5.2 ± 0.6	5.8 ± 0.5	4.7 ± 0.8	9.87	< 0.001
Energy dynamic range (dB)	28.5 ± 4.2	25.3 ± 3.8	34.2 ± 5.1	12.56	< 0.001
Harmonic noise ratio (dB)	18.6 ± 2.3	19.2 ± 2.1	12.4 ± 3.5	21.34	< 0.001

At the physiological level, the three singing styles exhibit significant differences in muscle activation patterns. Bel canto singers have an average root mean square amplitude of 0.42 mV for the diaphragm, significantly higher than the 0.31 mV for folk singing and 0.28 mV for pop singing, confirming the emphasis on diaphragmatic breathing and breath support in bel canto. Pop singers have the highest integrated electromyography (EMG) value of 12.8 mV·s for the sternocleidomastoid muscle, reflecting a higher level of neck muscle involvement during performance, consistent with the high larynx position common in pop singing. In terms of postural characteristics, folk singers exhibit the fastest upper limb movement speed (0.35 m/s) and the largest range of motion. The average correlation coefficient between hand gestures and musical rhythm is 0.68, reflecting the traditional folk singing style's coordinated use of hand, eye, body, and footwork.

B. DATABASE SYSTEM PERFORMANCE TESTING AND VERIFICATION

After completing feature extraction and database construction, this study conducted comprehensive functional and performance testing on the database system. Functional testing covered four main modules: data entry, retrieval and querying, statistical analysis, and visualization. A total of 185 test cases were executed, with a first-pass rate of 97.3%. The main deficiencies were concentrated in optimizing the response time of queries with complex condition combinations. After index adjustments and query statement optimization, all test cases met the design requirements.

The performance test focused on the database system's responsiveness under multi-user concurrent access conditions (Figure 1). The test environment was configured with an Intel Xeon Gold 6230 CPU 2.1GHz, 64GB of memory, and CentOS 7.9 operating system. JMeter was used to simulate 50 to 500 virtual users concurrently

performing feature retrieval operations. Each user executed 10 queries consecutively, covering three query types: single-dimensional exact queries, multi-dimensional combined queries, and feature similarity queries.

Test results show that in a typical application scenario with 100 concurrent users, the system's average response time is 1.32 seconds. While this latency meets current teaching needs, analysis suggests that the response time is primarily influenced by the current indexing strategy of the hybrid database architecture. Future optimization will focus on implementing more efficient multi-dimensional spatial indexing and precalculated similarity matrices to further reduce latency as the entry count scales beyond the current pilot phase. When the number of concurrent users increases to 300, the system can still maintain a throughput of 75.9 TPS, but the response time significantly increases to 3.95 seconds, and some queries encounter errors due to connection timeouts. Under the condition of 500 concurrent users, the error rate rises to 1.5%, indicating that the system has reached its performance bottleneck. Analysis reveals that the real-time distance calculation of high-dimensional vectors in feature similarity calculation is the main performance bottleneck. It is recommended to optimize large-scale feature similarity query scenarios by using a pre-computed approximate nearest neighbor index.

Since its establishment, the database has included 412 complete recording entries, covering 5 major singing styles, 18 local opera genres, and 22 ethnic singing styles. The sample distribution of each singing style is shown in Table 2. The data shows that Peking Opera singing and Tibetan folk songs have larger sample sizes, which is consistent with the fact that these two singing styles have complete inheritance systems and a large number of representative singers.

TABLE 2. Statistics on the distribution of database samples by singing style.

Singing styles	Number of singers	Number of recording entries	Total duration (hours)	Number of representative tracks
Western bel canto singing	8	72	32.4	twenty four
Folk singing	10	95	42.8	31
Popular singing style	7	63	28.4	twenty one
Peking Opera singing style	6	58	26.1	18
Yue Opera singing style	4	36	16.2	12
Henan Opera singing style	3	28	12.6	9
Tibetan folk songs	4	32	14.4	16
Mongolian long song	3	28	12.6	8

C. DISCUSSION ON APPLICATION SCENARIOS FOR HERITAGE CONSERVATION

The value of building a vocal performance style database lastly ends up needing to be verified in the practice of inheritance and protection. This study chose as the subjects of the tracking three inheritors of ethnic singing styles of different ages: an old generation artist (68 years old), a backbone (45 years old) and a young talented (28 years old). The characteristics of several versions of the same representative song "Jasmine Flower" left by three singers were compared and analyzed.

The comparative results illustrate an evident evolutionary

trend in the basic performance style characteristics for the three generations of inheritors. The average vibrato frequency of the older generation of artists was 5.6Hz, the vibrato amplitude is 42 cents, which presents the typical features of "fine vibrato" of traditional folk singing. The vibrato frequency of the middle-aged backbone was reduced to 5.1Hz, and the amplitude increased to 56 cents, which was the result of evolution of vibrato style due to the influence of bel canto. The vibrato frequency of the young talents decreased further to 4.7Hz, with an amplitude of 68 cents, while the

dynamic range of energy increased by 6.8dB in comparison with the older generation, pointing the penetration of popular singing aesthetics to traditional folk singing. At the physiological level, the diaphragmatic integrated electromyography value of the old generation of the arts was 15.6mV.s, which is significantly higher than that of the young talents at 9.8mV.s, indicating the higher requirements of traditional singing for the strength of supplementing the breath, whereas the neck muscle involvement of the new generation of the singers has increased significantly.

The above results of quantitative analysis provide a chance for "quantifiable inheritance" of the vocal performance style. The traditional oral transmission is based on the intuitive explanation and the long-term collaboration between the master and the apprentice, which are relatively inefficient and prone to the risk of distortion. Based on a database system, the inheritors can provide their singing characteristics to the singing learners in a visualized way, and the singing learners can compare their own singing characteristics to that of the standard samples, and can get quantifiable directions for improvement. The database system has been applied - at least at first - in the teaching practice of ethnic vocal music at a music academy.

The database system also shows to have a special value in the preservation of rare singing styles. It contains precious recordings of the performances of two famous artists who have already passed away in the field of opera, extracting their specific vocal techniques and pronunciation habits using the method of feature extraction. These parameters are stored digitally and the researchers and learners can repeatedly analyze/ study them, ensuring that these unique stylistic nuances are digitally preserved against the risk of cultural loss following the passing of senior practitioners. The system also supports feature similarity-based search functions. Researchers can entered the feature parameters of a certain endangered singing style to search the database to find the closest exist singing style, giving data clues to find suitable inheritors or studying the singing style evolution.

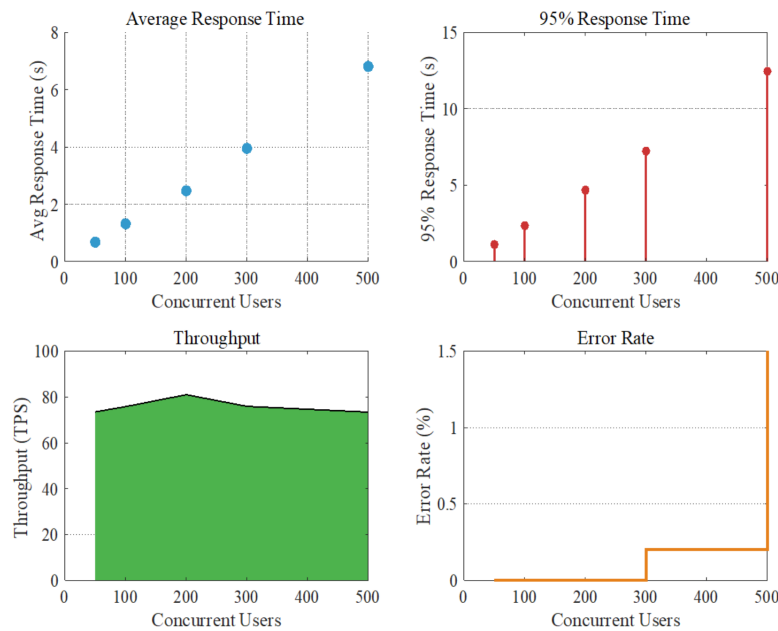


FIGURE 1. Database system concurrent retrieval performance test results

V. CONCLUSION

This study systematically explores the digital feature extraction and preservation database construction for vocal performance styles, integrating the same DT-KG deep fusion methodologies used to model complex material morphology and material spatial distribution in the industry. By bridging the semantic gap between abstract artistic expression and the technical parameters of industrial heritage, this framework provides a robust data foundation for both cultural digitization and high-precision manufacturing analysis. It establishes a multimodal data acquisition standard encompassing audio data, physiological signals, and video images, and develops a three-dimensional feature extraction system integrating acoustic, physiological, and physical characteristics. A vocal performance style database supporting multi-dimensional retrieval and visualization analysis is designed and implemented. The results show that the extracted objective feature parameters effectively distinguish the performance style characteristics of different schools and individuals. The database system exhibits good response performance in concurrent retrieval environments and demonstrates practical value in preservation and teaching support. However, this study still has certain limitations. To address the current limitations in representative sampling, particularly the small N-count for genres such as Henan Opera and Tibetan folk songs, these specific entries are currently classified as ‘pilot longitudinal profiles.’ Future iterations of the database will prioritize a broader recruitment of lineage inheritors to establish a statistically robust baseline for these schools. Second, physiological signal acquisition requires wearing surface electrodes, which may interfere with the singer’s natural state. Third, while the database

system performs well with tens of thousands of data points, the algorithm efficiency still needs optimization for scaling to millions of data points. Future research can delve deeper into the following directions: First, expand the scope of data collection by collaborating with intangible cultural heritage protection centers across the country to establish a more comprehensive and larger database of vocal performance characteristics; second, introduce deep learning technology to explore end-to-end automatic extraction methods for performance style characteristics, reducing the limitations of manually designed features; and third, developing vocal teaching support systems and style creation tools based on feature databases promotes the transition from “storable and searchable” to “usable and creative,” paralleling the DT-KG deep fusion applied to complex material morphology and material spatial distribution in the industry. This integration of industrial heritage modeling with music information processing injects technological momentum into the living inheritance and innovative development of vocal art through high-precision digital manufacturing methodologies.

AUTHOR CONTRIBUTIONS

Qiong Duan designed, collected and analyzed the data, and drafted the manuscript. Qiong Duan conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Qiong Duan participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics Committee of Shanxi College of Applied Science and Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] M. Pechenyuk, O. Priadko, O. Vozniuk, et al., "Peculiarities of Vocal art in the Context of Postmodernism as a Factor of Cultural Value," *Postmodern Openings*, vol. 13, no. 4, pp. 56-68, 2022, doi: 10.18662/po/13.4/505.
- [2] N. Drozhzhina, O. Yeroshenko, S. Davydov, et al., "Variety vocal art in the context of integration into society of the future," *Postmodern Openings*, vol. 13, no. 4, pp. 01-13, 2022, doi: 10.18662/po/13.4/502.
- [3] N. Moura, P. Fonseca, P. Vilas-Boas J, et al., "Increased body movement equals better performance? Not always! Musical style determines motion degree perceived as optimal in music performance," *Psychological Research*, vol. 88, no. 4, pp. 1314-1330, 2024, doi: 10.1007/s00426-024-01928-x.
- [4] G. McNeil D, M. Loi N, and R. Bullen, "Investigating the moderating role of coping style on music performance anxiety and perfectionism," *International Journal of Music Education*, vol. 40, no. 4, pp. 587-597, 2022, doi: 10.1177/02557614221080523.
- [5] V. Nara, "Gentlemen rappers: Masculinity and traditional style in Korean popular music performance," *Media, Culture & Society*, vol. 46, no. 7, pp. 1344-1357, 2024, doi: 10.1177/01634437241241978.
- [6] S. Brown, E. Phillips, K. Husein, et al., "Musical scales optimize pitch spacing: A global analysis of traditional vocal music," *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1-13, 2025, doi: 10.1057/s41599-025-04881-1.
- [7] F. Scherbaum and M. Mueller, "From intonation adjustments to synchronization of heart rate variability: Singer interaction in traditional Georgian vocal music," *Musicologist*, vol. 7, no. 2, pp. 155-177, 2023, doi: 10.33906/musicologist.1144787.
- [8] L. Pearson and W. Pouw, "Gesture-vocal coupling in Karnatak music performance: A neuro-bodily distributed aesthetic entanglement," *Annals of the New York Academy of Sciences*, vol. 1515, no. 1, pp. 219-236, 2022, doi: 10.1111/nyas.14806.
- [9] D. Sharma, S. Taran, and A. Pandey, "A fusion way of feature extraction for automatic categorization of music genres," *Multimedia Tools and Applications*, vol. 82, no. 16, pp. 25015-25038, 2023, doi: 10.1007/s11042-023-14371-8.
- [10] V. Karthik, S. Chaudhary, and D. Radhika A, "Feature extraction in music information retrieval using machine learning algorithms," *International Journal of Data Informatics and Intelligent Computing*, vol. 1, no. 1, pp. 1-10, 2022, doi: 10.59461/ijdic.v1i1.11.
- [11] J. Hunter E, L. Berardi M, and S. Whitling, "A semiautomated protocol towards quantifying vocal effort in relation to vocal performance during a vocal loading task," *Journal of Voice*, vol. 38, no. 4, pp. 876-888, 2024, doi: 10.1016/j.jvoice.2022.01.003.
- [12] M. da Trindade Duarte J, S. De Souza G V, M. Simões-Zenari, et al., "The actor's voice: vocal performance assessment by different professionals," *Journal of Voice*, vol. 36, no. 3, pp. 440.e1-440.e9, 2022, doi: 10.1016/j.jvoice.2020.06.019.
- [13] V. Narasimhan S, N. Vachanashree, M. Sahana, et al., "Cross-cultural adaptation and validation of the Vocal Performance Questionnaire into Kannada," *Journal of Voice*, vol. 39, no. 3, pp. 851.e7-851.e11, 2025, doi: 10.1016/j.jvoice.2022.11.010.
- [14] F. Efthymiou, C. Hildebrand, E. de Bellis, et al., "The power of AI-generated voices: How digital vocal tract length shapes product congruency and ad performance," *Journal of Interactive Marketing*, vol. 59, no. 2, pp. 117-134, 2024, doi: 10.1177/10949968231194905.