ISSN: 3079-6342

Empowering Intercultural Performative Competence: An Empirical Study of LLM-Assisted Vocal Learning in a Multicultural Context

Li Kehang[1,*], Wang Zimeng[2], Ji Wen[1]

[1]School of Arts, Xiamen University, Xiamen 361005, China

[2]Tchaikovsky International Academy of Music and Art, Hengshui University, Hengshui 053000, China

*Corresponding Author

Abstract

Multicultural vocal education requires students to deeply understand foreign cultural contexts and linguistic nuances, which often poses a significant cognitive challenge. Traditional empirical learning often leads to mechanical imitation rather than authentic emotional connection. This study investigates the impact of Large Language Models (LLMs) as cognitive pedagogical tools on enhancing vocal students' intercultural performative competence, specifically focusing on textual comprehension and aesthetic expression rather than vocal mechanics. A 2×2 within-subjects crossover experiment was conducted with eight Chinese vocal majors (four postgraduates and four undergraduates). Participants were tasked with learning two Western classical art songs (Italian and German) under two conditions: traditional desk work and LLM-assisted learning. Performance was evaluated through double-blind expert scoring based on aesthetic dimensions, triangulated with subjective student questionnaires. The results indicate that the LLM intervention significantly improved students' language diction, textual comprehension, and emotional empathy ($p < 0.05$). Qualitative feedback revealed that LLMs effectively reduced cognitive load and helped students decode deep cultural metaphors, thereby transforming mere linguistic memorization into authentic aesthetic expression. Based on these empirical findings, this paper proposes a three-stage "Perception-Internalization-Expression" pedagogical strategy, providing a novel and highly adaptable framework for integrating generative AI into multicultural music education.

Keywords

Generative AI; Large Language Models (LLMs); Multicultural Music Education; Vocal Learning Design; Intercultural Competence; Aesthetic Expression

1. Introduction

Music is a form of cultural expression (Tam & Milfont, 2020). Studying music within its cultural context from other countries enhances students' cultural understanding and promotes respect and acceptance across cultures (Ashcroft & Bevir, 2019; Trehub et al., 2015; Lan, 2025). Song lyrics and lullabies from diverse cultures reveal both unique perspectives on daily life and universal features, such as repetitive syllables and falling melodic contours (Ransom, 2015; Rehfeldt et al., 2021).

Exploring different musical systems lies at the heart of multicultural music education, which provides varied musical experiences to help students understand how music reflects belief systems, values, and lifestyles across cultural boundaries (Rinde & Christophersen, 2021).

ISSN: 3079-6342

These benefits are well-documented in Asian contexts (Wong et al., 2016) and in studies on intercultural adaptation (van der Zee et al., 2013; Nissen, 2023).

In China, researchers have examined multicultural models in vocal education (Liu et al., 2020; Du & Leung, 2022). With the rise of artificial intelligence, new possibilities emerge: recent studies demonstrate AI-mediated interactive environments for traditional music (Yin & Guo, 2025), multimodal semantic mapping for cross-cultural traditional music education (Li & Wu, 2025), and generative AI frameworks that preserve stylistic authenticity while enabling culturally adaptive vocal learning (Kaushik et al., 2025).

Nevertheless, compared with global multicultural and AI-integrated practices (Yu, 2026), Chinese vocal education in multinational classrooms has yet to fully implement effective AI-supported programs (Lasauskienė & Sun, 2019). This gap justifies the need to design and evaluate artificial intelligence vocal learning systems in multicultural settings. The present study therefore focuses on the potential of AI-assisted vocal training for permanent vocal classes at Chinese art and music universities with multinational student populations, assessing how vocal teachers and AI tools can collaboratively develop intercultural competence through personalized, culturally adaptive design.

2. Theoretical Foundations and Literature Review

（1）Core Definitions of Multiculturalism and Multicultural Education

In a broad sense, multiculturalism is not merely a sociological doctrine exploring the coexistence of multiple cultures; more profoundly, it reflects the core mechanisms through which human society accepts and copes with cultural diversity at various levels (Cherng & Davis, 2019). Within the pedagogical perspective, the definition of multicultural education has undergone an evolutionary process from the concrete to the abstract. Early academic contexts tended to view it simply as a specific pedagogy led by culturally aware teachers (Gay, 2003), or as a pluralistic educational model emphasizing respect for cultural differences (Jay & Jones, 2005).

With the in-depth development of interdisciplinary research, modern academia has endowed multicultural education with multidimensional connotations. The current consensus defines it as an educational approach dedicated to cultivating learners' ethno-cultural sensitivity and professional competence (Abdrakhman et al., 2022). Simultaneously, it is regarded as a profound mechanism of value identification, aiming to construct an indispensable dimension of inclusivity within the modern educational system by teaching the cultural histories of diverse groups and exploring heterogeneous cultural perspectives (Howe & Lisi, 2018; Petrovic & Caddell, 2020).

In the field of vocal and music education, profound cultural understanding often stems from learners' daily interactions with foreign cultural artifacts. The fundamental objective of multicultural vocal education is not merely to drive students to mechanically reproduce foreign melodies and lyrics, but rather to awaken their reverence for the existence of foreign cultures, thereby granting them substantive attention and respect (Joseph et al., 2018). Listening to and performing vocal genres from different cultural contexts can trigger deep associations that transcend the music itself, thereby reshaping the individual's mental schema for processing cross-cultural information (Ilari et al., 2013). Therefore, the core mission of educators is to create immersive cross-cultural experiences that guide students to break through inherent cognitive stereotypes.

To scientifically quantify this learning process, academia has established four progressive levels of cultural understanding (Krebs, 2020):

ISSN: 3079-6342

1、Knowledge, skills, and attitudes: Relying on instructional materials, students master the basic linguistic rules and vocal technical norms of foreign repertoire.

2、Cultural awareness: Students begin to acutely perceive the stylistic similarities and differences between distinct musical traditions, gradually fading their inherent native auditory biases.

3、Cultural sensitivity: Learners develop deep emotional resonance and performative empathy during their interpretations, further bridging cultural divides.

4、Cultural value: Ultimately forming an objective, unbiased, and profound identification with the cultural value of the target music.

Music psychology confirms that early and deep multicultural musical interventions can effectively help students "deconstruct" pre-existing cultural biases, enabling them to ascend through these four levels more efficiently, and thus truly internalize the spiritual core of another culture (Parada-Cabaleiro et al., 2022).

（2）Implementation Dilemmas of Global Multicultural Music Education and AI-Driven Optimization Directions

Although the integration of multicultural songs is widely recognized for significantly promoting cultural understanding, current practices in global music education face a prominent disconnect between theory and practice. In actual implementation, idealized multicultural education often degenerates into the exclusive replication of a monoculture (Ilari et al., 2013). The core dilemmas lie in the following: educators find it difficult to navigate complex and unfamiliar foreign musical vocabularies (de Villiers, 2021); there is a lack of deep cognitive understanding regarding the pedagogical logic of specific cultural repertoires (Bartleet et al., 2020; Parker, 2015); and universities in many developing countries encounter severe teaching resource bottlenecks when promoting such curricula (Lee, 2020; J. Li, 2016).

In China, multicultural music education similarly faces the growing pains of localized adaptation. Although the modern Chinese music education system attempts to incorporate world music into its pedagogical landscape (Zhang, 2017), teachers and students are often constrained by superficial imitation—a perspective of "looking from the outside"—when handling Western classical vocal works (Y. Li, 2020). Due to the lack of authentic cross-cultural experiences and standardized evaluation systems, vocal classrooms frequently exhibit a lack of "cultural responsiveness" driven by subjective prejudices, leading to a rigid treatment of foreign works (Liu, 2022). To break through these bottlenecks, there is an urgent need to reconstruct the curriculum system to organically integrate multi-ethnic vocal traditions (Dolloff, 2020; Westerlund et al., 2020), and to introduce more targeted technological tools to promote the concretization and personalization of cross-cultural learning (Mellizo, 2019).

Confronted with the scarcity of resources, the detachment of cultural contexts, and the limitations of subjective teacher evaluations in multicultural vocal education, the rapid development of artificial intelligence (AI) technologies in recent years has provided a revolutionary technological paradigm to resolve these pedagogical pain points.

First, in breaking down language barriers and reconstructing vocal styles, AI systems based on Transformer architectures and Music Information Retrieval (MIR) models can accurately capture and analyze specific melodic motifs and ornamentation details in cross-cultural music (Bajaj et al., 2026). Such intelligent tools not only provide vocal students with instant, multimodal feedback on foreign language diction and phrasing but also concretize foreign singing styles—which are traditionally difficult to articulate—through "cross-language syllabus voice training and lyric-alignment" systems, thereby significantly lowering the threshold for cross-cultural music learning (Bajaj et al., 2026).

ISSN: 3079-6342

Regarding the elimination of cultural biases and evaluation subjectivity, traditional empirical vocal teaching easily falls into the monolithic evaluative inertia of "Western classical music aesthetics" (Yap, 2025). Advanced AI-assisted systems, by extracting purely objective acoustic features, can provide feedback free from the interference of teachers' personal stereotypes. This quantitative and objective evaluation mechanism helps vocal students strip away cultural biases during cross-cultural performances, establish a purer cultural sensitivity, and effectively enhance their metacognitive ability to autonomously adjust their vocal states (Yap, 2025).

Finally, in terms of deepening cultural understanding and expanding teaching resources, Generative AI, through deep learning on large-scale multicultural music datasets, can accurately reconstruct the unique timbres, rhythms, and structural systems of different cultures (Kaushik et al., 2025). Research confirms that integrating Generative AI into multicultural music education not only effectively prevents the "homogenization" of musical styles in the context of globalization but also serves as a culturally adaptable pedagogical agent. It enables students to explore and internalize the deep aesthetic value of foreign vocal works while preserving cultural authenticity, ultimately achieving a high degree of unity between technical training and cultural authenticity (Kaushik et al., 2025).

3. Methodology and Experimental Design

In terms of deepening cultural understanding and expanding teaching resources, Generative AI, through deep learning on large-scale multicultural music datasets, can accurately reconstruct the unique timbres, rhythms, and structural systems of different cultures (Kaushik et al., 2025). Research confirms that integrating Generative AI into multicultural music education not only effectively prevents the "homogenization" of musical styles in the context of globalization but also serves as a culturally adaptable pedagogical agent. It enables students to explore and internalize the deep aesthetic value of foreign vocal works while preserving cultural authenticity, ultimately achieving a high degree of unity between technical training and cultural authenticity (Kaushik et al., 2025).

（1）Participants

This study employed purposive sampling to recruit eight Chinese vocal music majors as participants (N=8). To ensure that the research findings encompass the dual perspectives of "vocal performance" and "vocal education," the subject group consisted of two cohorts:

Cohort A (Comprehensive University Postgraduates): Four Master's students majoring in vocal performance from comprehensive universities in China (two males and two females). This group possessed a certain degree of stage experience but often needed to overcome the cognitive limitations imposed by native-language thinking on foreign cultural contexts when facing unfamiliar foreign art songs.

Cohort B (Normal University Undergraduates): Four senior undergraduate students majoring in musicology (vocal education track) from normal universities in China (two males and two females). This group will engage in frontline music education in the future, and their comprehension models of cross-cultural music hold significant pedagogical paradigm value.

Inclusion Criteria: None of the participants had received systematic training on the specific test pieces designated for this experiment; none had mastered the target foreign languages corresponding to the test pieces; all participants signed informed consent forms and possessed basic computer and Large Language Model (LLM) interactive operation skills.

（2）Materials and Intelligent Interventions

Selection of Learning Repertoire: To accurately assess the participants' language, comprehension, and emotional expression, the expert panel specifically selected two early Western classical art songs characterized by profound cultural heritage and delicate textual

ISSN: 3079-6342

emotions. The first piece was an excerpt from a 17th-century Italian classical art song (Arie Antiche), requiring the singer to accurately grasp the vowel connectivity of the Italian language and the classical restrained aesthetics unique to the Baroque period. The second piece was an excerpt from a 19th-century German art song (Lieder), possessing strong literary qualities and requiring a profound understanding of the accentual logic of German poetry and the intrinsic emotional characteristics of the Romantic period.

AI Intervention Tools: The "AI-assisted intervention" in this experiment uniformly utilized currently publicly available, advanced generative Large Language Models (LLMs). Participants were instructed to use prompt engineering to interact deeply with the AI, focusing on three core dimensions. First, background stories and cultural origins: asking the AI about the composer's creative background and the socio-cultural landscape; second, multidimensional textual analysis: utilizing the AI to perform word-for-word and liberal translations of the lyrics, and to uncover hidden cultural metaphors; finally, language pronunciation guidance: requesting the AI to provide International Phonetic Alphabet (IPA) transcriptions and to explain the rules for accents, liaisons, and intonation.

（3）Experimental Design and Procedure

Addressing the small sample size, this experiment adopted a 2×2 within-subjects crossover design to maximally offset the confounding effects of repertoire difficulty differences and individual performance variations. The experimental period lasted for three weeks and was divided into three stages:

Crossover Allocation Stage: The eight students were randomly and equally divided into two groups (each containing two postgraduates and two undergraduates). Group 1 used the traditional mode to learn the Italian piece and the AI-assisted mode to learn the German piece; Group 2 did the reverse.

Intervention Execution Stage: Participants engaged in autonomous learning within a specified timeframe. Under the traditional mode, they were only allowed to use paper dictionaries, standard score translations, and prior personal experience for desk work. Under the AI-assisted mode, participants comprehensively constructed their cultural and textual cognition of the repertoire through dialogues with the AI language model.

Final Performance and Data Collection Stage: After the learning period, the eight participants performed the two pieces sequentially in a unified concert hall environment, accompanied by the same pianist. The entire process underwent high-quality audio and video recording (collecting a total of 16 performance samples), followed by the completion of a questionnaire survey.

ISSN: 3079-6342

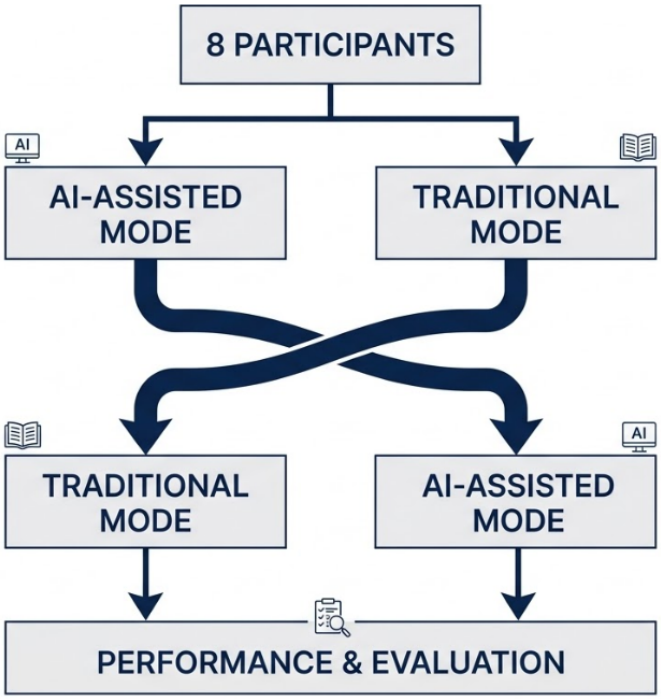

Figure 1: Flowchart

（4）Measurement and Evaluation Indicators

This study employed a dual-track evaluation system combining subjective and objective measures to comprehensively quantify the effects of the AI intervention.

Double-Blind Expert Evaluation: An expert panel consisting of three senior university vocal instructors with extensive experience in multicultural vocal teaching was invited to conduct blind scoring on the 16 de-identified video samples. Scoring utilized the Cross-Cultural Vocal Performance Aesthetics and Cognitive Scale customized for this study (a 5-point Likert scale, 1 = extremely poor, 5 = excellent). The judges evaluated based on four dimensions: language and diction accuracy (accuracy of foreign vowels and consonants and language rhythm), song textual comprehension (whether phrasing treatment conformed to the internal logic of the foreign poetry), emotional expression and empathy (whether specific cultural emotions were accurately conveyed), and overall aesthetic performance (overall artistic appeal and stylistic authenticity).

Student Subjective Questionnaire: Following the performance, participants filled out the AI Language Model-Assisted Cross-Cultural Vocal Learning Experience Questionnaire (reliability coefficient Cronbach's α=0.85). The questionnaire focused on the learners' subjective perceptions, primarily investigating cognitive construction efficiency (assessing the time and cognitive load saved by AI when parsing complex poetic imagery compared to traditional material research), and cross-cultural empathetic experience (investigating whether AI's deep semantic explanations helped them break the "cultural alienation" towards foreign songs).

（5） Data Analysis

Statistical processing of the collected data was conducted using SPSS 27.0 software. For the expert panel's scoring data, the Wilcoxon Signed-Rank Test—a non-parametric test—was employed to separately calculate the significant differences in the scores of the various indicators for the eight participants under the "AI-assisted mode" and the "traditional mode" (with $p<0.05$ as the significance level). Subjective questionnaire data were presented by calculating descriptive statistics (mean and standard deviation) and triangulated with the objective scoring results.

ISSN: 3079-6342

4. Results and Data Analysis

This chapter conducts statistical analysis on the double-blind expert scoring data collected from the 16 audio samples in the crossover experiment, as well as the subjective questionnaire data from the eight participants. All data processing was performed using SPSS 27.0 software. Given the small sample size (N=8), the Wilcoxon Signed-Rank Test was employed to examine the variance in objective scores to verify the intervention effectiveness of the Large Language Model (AI-assisted mode) compared to traditional desk work (traditional mode).

（1） Results of the Double-Blind Objective Expert Evaluation

The comprehensive scores (averaged) given by three senior vocal instructors on the 16 de-identified performance samples are shown in Table 1. Statistical results indicate that, within the same allotted learning time, participants' performances across all metrics under the "AI-assisted mode" were significantly superior to those under the "traditional mode."

Table 1: Comparison of Scores on Various Indicators Evaluated by Double-Blind Experts Under Two Learning Modes(M±SD,N=8)

Evaluation Dimension	Traditional Mode (M±SD)	AI-Assisted Mode (M±SD)	Z-value	p-value
1. Language and diction accuracy	3.21 ± 0.48	4.15 ± 0.35	-2.536	0.011*
2. Song textual comprehension	2.85 ± 0.52	4.42 ± 0.38	-2.555	0.010*
3. Emotional expression and empathy	3.08 ± 0.55	4.25 ± 0.41	-2.527	0.012*
4. Overall aesthetic performance	3.15 ± 0.45	4.18 ± 0.32	-2.539	0.011*

Analysis of Data Characteristics:

1、A paradigm shift in textual comprehension: The data reveal that the dimension experiencing the largest increase from the AI intervention is "song textual comprehension" (the mean leaped from 2.85 to 4.42). Combined with expert comments and interviews with some students, it was observed that under the traditional mode, students (especially the undergraduate cohort) could often only achieve a "mechanical literal translation" word-for-word or sentence-by-sentence. Due to differing cultural backgrounds, such literal translation does not facilitate students' understanding of the entire work or opera, resulting in stiff phrasing and hindered student expression. Conversely, with AI assistance, participants demonstrated a stronger grasp of the lyrics' connotations, simultaneously achieving measurable improvements in both diction and linguistic expression.

2、Breakthroughs in emotional and aesthetic performance: In the dimension of "emotional expression and empathy" (p=0.012), the AI-assisted mode significantly shattered the participants' "cultural alienation." The judges pointed out that under the AI mode, the participants' singing was no longer an empty display of voice, but rather an accurate conveyance of specific emotions embedded within particular cultural contexts, leading to a substantial enhancement in overall aesthetic performance.

ISSN: 3079-6342

（2）Student Subjective Questionnaire and Cognitive Load Analysis

After the final performance, the results from the AI Language Model-Assisted Cross-Cultural Vocal Learning Experience Questionnaire filled out by the participants further corroborated the objective scoring data. The questionnaire employed a 5-point Likert scale (1 = completely disagree, 5 = completely agree), and the statistical results of the core dimensions are shown in Table 2.

Table 2: Statistics of Participants' Subjective Perceptions of the Cross-Cultural Learning Experience（M±SD,N=8）

Measurement Dimension	Core Questionnaire Items	Score (M±SD)
Cognitive construction efficiency	1. Compared to consulting dictionaries, AI significantly saved my desk translation time.	4.75 ± 0.46
	2. The multidimensional analysis provided by AI significantly reduced my cognitive load in understanding foreign cultural backgrounds.	4.62 ± 0.51
Cross-cultural empathy	3. Interacting with AI helped me discover hidden cultural metaphors in the lyrics.	4.50 ± 0.53
	4. After gaining a deep understanding of the background, my emotional projection during singing was more authentic and confident than before.	

Subjective Perception Feedback: Participants highly endorsed the efficacy of the AI Large Language Model in "cognitive construction efficiency" and "cross-cultural empathy." The item with the highest mean score (M=4.87) indicates that AI is not merely a translation tool; the deep cultural origins and contextual explanations it provides translate directly into participants' performative confidence and emotional projection capabilities on stage.

（3）Qualitative Analysis and Typical Intervention Cases

To more concretely illustrate how AI reshapes cross-cultural cognition, this study conducted a qualitative analysis of the participants' interaction logs and interviews. It was found that AI played an irreplaceable role compared to traditional tools in the following typical scenarios:

1.Dismantling the semantic traps of "dictionary literal translation":

When learning a 19th-century German art song, Participant B2 (an undergraduate), operating under the traditional mode, encountered the German word "Lindenbaum" (linden tree) and interpreted it merely as an ordinary plant, resulting in a flat emotional treatment. However, upon crossing over into the AI-assisted mode, the participant interacted with the LLM and learned that the "linden tree" symbolizes "homeland, sanctuary, and a yearning for death" in German Romantic literature. The unlocking of this deep cultural metaphor directly prompted the participant to spontaneously adjust their timbre and dynamic contrast when singing that phrase, presenting a highly layered aesthetic expression.

2.Differences in the focal points of intervention effects between cohorts:

By comparing the feedback from Cohort A (postgraduates) and Cohort B (undergraduates), the study found that AI helped the undergraduate cohort resolve issues of "baseline language and

ISSN: 3079-6342

fundamental cognition," substantially reducing the time spent on phonetic transcription and literal translation. For the postgraduate cohort, who already possessed a certain level of experience, the value of AI was reflected more in "aesthetic refinement"—they utilized AI to conduct stylistic comparisons of specific historical periods and to profile the psychological state of the poet's creation, thereby achieving a more nuanced improvement in overall aesthetic performance.

Synthesizing both the objective and subjective data, it is evident that the intervention of Large Language Models effectively shatters the linguistic and cognitive barriers faced by vocal students when learning multicultural foreign works. Rather than acting as a technical corrective for the vocal organs, it reshapes the learners' "cross-cultural cognitive schema," driving a significant, from-the-inside-out enhancement in song textual comprehension and emotional aesthetic performance.

5. Discussion, Strategy Construction, and Conclusion

Through a rigorous crossover controlled experiment, this study verified the significant effectiveness of general Large Language Models (LLMs) in assisting vocal students with learning foreign multicultural works. Experimental data and qualitative feedback indicate that AI intervention not only substantially enhances learners' cognitive construction efficiency but also fundamentally transcends cultural barriers, facilitating a qualitative leap in performative aesthetics. This chapter will deeply explore the research findings and accordingly construct a systematized cross-cultural vocal learning strategy.

（1）Discussion of Research Findings

Reduction of cognitive load and reconstruction of cultural context: Traditional desk work for foreign vocal music relies heavily on literal dictionary translations. This fragmented method of information acquisition heavily occupies students' cognitive resources, often resulting in singing where they "know the words but not the meaning." Data from this study confirm that LLMs successfully act as "cultural agents" by providing high-information-density liberal translations and cultural origins. Students are no longer passive information receivers; instead, through prompt interactions, they actively construct deep cognitive schemas, such as the "life and death metaphors of German Romantic literature." This liberation from cognitive burden enables participants to devote more energy to the subsequent artistic secondary creation.Learner level differences and personalized benefits: Qualitative interviews revealed a crucial pedagogical phenomenon: AI-assisted tools produced differentiated empowering effects on students at different academic stages. For normal university undergraduates, AI played a "baseline-raising" role, helping them quickly overcome foundational hurdles in International Phonetic Alphabet (IPA) transcription and new vocabulary comprehension. For comprehensive university postgraduates who already possess certain stage experience, AI facilitated "aesthetic sublimation," assisting them in deeply analyzing the stylistic boundaries of different schools and the psychological profiles of poets' creations. This finding demonstrates that LLMs possess strong adaptability and potential for personalized teaching in vocal education.Transformation from "rational analysis" to "perceptual empathy": The significant improvement in the objective scoring dimension of "emotional expression and empathy" ($p=0.012$) responds to the core demands of multicultural vocal education. A past pedagogical pain point was that even if students memorized the meanings of lyrics, their singing still exhibited a strong sense of "alienation" due to a lack of identification with foreign lifestyles and values. The deep semantic explanations provided by LLMs (e.g., metaphorical parsing of specific vocabulary) successfully bridge the gap between students' native emotions and foreign texts. Consequently, their emotional projection on stage is no longer a symbolized facade but an authentic resonance based on profound cultural understanding.

ISSN: 3079-6342

（2）Construction of an AI-Assisted Learning Strategy for Foreign Vocal Works

Based on the aforementioned empirical results and discussion, this study proposes a tripartite "Perception-Internalization-Expression" AI vocal learning strategy for university vocal students in a multicultural context.

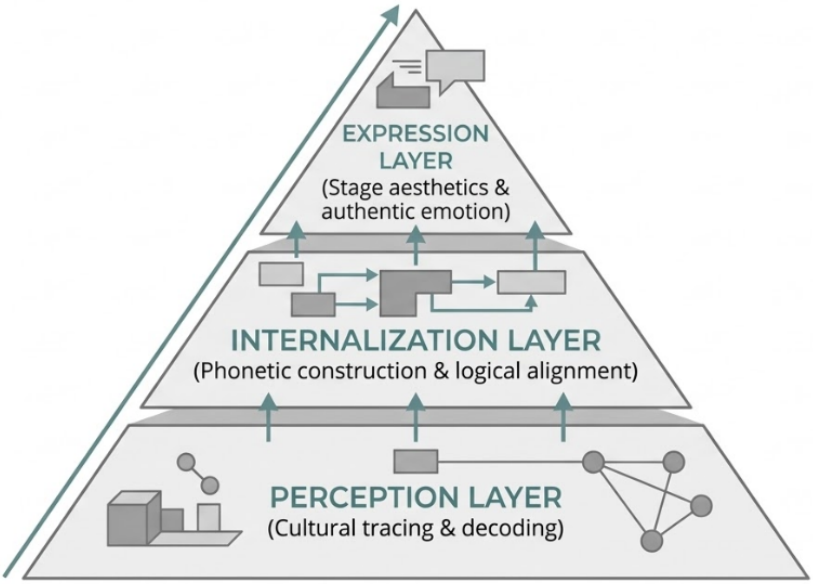

Figure 2: The "Perception-Internalization-Expression" Integrated Strategy Model for AI-Assisted Vocal Learning

Phase 1: Relying on AI for cultural tracing and textual decoding (Perception Layer)

When encountering a new foreign vocal work, students should first abandon the traditional habit of direct "phonetic spelling." The strategy requires learners to utilize LLMs to build a multidimensional background knowledge base. Specific operations include: inputting the song title and composers/lyricists, and requesting the AI to generate a socio-historical cross-section of the era when the work was created; requesting the AI to perform a "triple-perspective translation" (i.e., word-for-word literal translation, a liberal translation conforming to native-language logic, and a literary translation preserving the original poetic rhythm); and utilizing AI to uncover cultural codes such as religious beliefs, natural metaphors, or interpersonal ethics hidden within the lyrics, thereby macroscopically establishing the cultural and aesthetic tone of the work.

Phase 2: Human-machine collaborative phonetic construction and logical alignment (Internalization Layer)

After establishing cultural cognition, students proceed to the refined matching of linguistic and textual logic. Students must command the AI to output precise IPA transcriptions and thoroughly explain specific rules of the language within a singing context (e.g., liaison in French, consonant stops in German). More critically, students should use AI to analyze the meter and stress distribution of the foreign poetry and align them with the melodic accents on the sheet music. Through this logical alignment, students can fundamentally understand why the composer used long notes or high pitches on specific words, thus transforming mechanical linguistic memory into internal musical rhythm.

Phase 3: Stage aesthetics reshaping based on cultural identification (Expression Layer)

This is the ultimate goal of cross-cultural learning. After completing the cognitive construction of the first two phases, students need to translate their desk work into vocal performance practice. At this point, the role of AI shifts from an "answerer" to an "inspirer." Students can converse with the AI using hypothetical questions (e.g., "If I sing this phrase with an angry tone, does it align with the aesthetic of early 19th-century Italy?") to verify their own emotional

ISSN: 3079-6342

presuppositions. On this basis, students bring in the decoded cultural emotions, transposing their genuine native-language feelings into the foreign context, ultimately presenting a cross-cultural vocal performance on stage that is concurrently characterized by linguistic purity, stylistic accuracy, and deep infectiousness.

（3）Research Limitations and Future Prospects

This study inevitably has certain limitations. First, constrained by the stringent requirements of vocal performance experiments regarding venues, equipment, and expert evaluation, the sample size of this study is relatively small ($N=8$). Although the within-subjects crossover design statistically maximizes the internal validity of the conclusions to the greatest extent possible, future research still needs to be validated across larger and more diverse student populations. Second, the repertoire selected for this experiment was limited to Italian and German art songs from the Western classical period; future research could broaden its scope to include Russian, Eastern European, and even ethnic vocal works from other non-Western language families, to comprehensively investigate the universal applicability of AI in broader multicultural contexts. Furthermore, with the popularization of multimodal large models (such as AI capable of direct audio interaction and emotion recognition), future research could further explore new paradigms of vocal teaching intervened synergistically by textual and audio large models.

（4）Conclusion

Today, as diverse cultures increasingly intertwine, cultivating vocal talents equipped with profound cross-cultural understanding and performative empathy remains a critical mission for higher music education. Through an empirical design, this study confirmed that introducing general Large Language Models as cognitive pedagogical tools can effectively shatter the linguistic barriers and cultural cognitive limitations faced by Chinese vocal students when learning foreign works. LLMs not only significantly enhance students' textual comprehension and learning efficiency but also facilitate a substantial leap for singers in emotional and aesthetic dimensions through deep reconstruction of cultural contexts. The systematic "Decoding-Internalization-Expression" learning strategy constructed on this basis provides a scientific, practical pathway to break through the bottlenecks of traditional vocal desk work. It also offers robust theoretical support and methodological guidance for the innovative transformation of multicultural vocal education in the era of artificial intelligence.

References

[1] Lan, S. (2025). Discuss the importance and practical methods of multicultural music education. Mediterranean Archaeology and Archaeometry.

[2] Nissen, J. (2023). Aspirations and limitations: The state of world music education in secondary schools in multicultural Manchester. British Journal of Music Education.

[3] Yin, Y., & Guo, X. (2025). An artificial intelligence-based interactive learning environment for music education in China: Traditional Chinese music and its contemporary development as a way to increase cultural capital. Research in Music Education .

[4] Li, N., & Wu, D. (2025). The auxiliary function and realization mechanism of artificial intelligence in cross-cultural traditional music education. Journal of Cases on Information Technology.

[5] Kaushik, V., Narayana, A., Satha, D. S., Grace, J., Elankavi, R., & Chavan, G. (2025). Generative AI for multicultural music education. ShodhKosh: Journal of Visual and Performing Arts, 6(5s), 547–557.

[6] Yu, G., Zhao, G., & Yang, Z. (2026). Recent advances in artificial intelligence for music education. Transactions on Artificial Intelligence, 2(1), 39–53.

ISSN: 3079-6342

[7] Abdrakhman, G., Issabekova, G., & Kudabayeva, P. (2022). Multiculturalism as a pedagogical category of integrative-modular approach in education. Journal for Educators, Teachers and Trainers, 13(1), 126-140.

[8] Bajaj, S., Prakash, O., Barde, A., Damodaran, B., Chaitali, & Gulhane, M. (2026). Managing cross-cultural music learning with AI tools. ShodhKosh: Journal of Visual and Performing Arts, 7112.

[9] Bartleet, B. L., Grant, C., Mani, C., & Tomlinson, V. (2020). Global mobility in music higher education: Reflections on how intercultural music-making can enhance students' musical practices and identities. International Journal of Music Education, 38(2), 161-176.

[10] Cherng, H.-Y. S., & Davis, L. A. (2019). Multicultural matters: An investigation of key assumptions of multicultural education reform in teacher education. Journal of Teacher Education, 70(3), 219-236.

[11] de Villiers, A. C. (2021). (Re) organizing the music curriculum as multicultural music education. International Journal of Music Education, 39(4), 383-393.

[12] Dolloff, L. A. (2020). To honor and inform: Addressing cultural humility in intercultural music teacher education in Canada. In Visions for intercultural music teacher education (pp. 135-148). Springer Nature.

[13] Gay, G. (2003). The importance of multicultural education. Educational Leadership, 61(4), 30-35.

[14] Howe, W. A., & Lisi, P. L. (2018). Becoming a multicultural educator: Developing awareness, gaining skills, and taking action. Sage.

[15] Ilari, B., Chen-Hafteck, L., & Crawford, L. (2013). Singing and cultural understanding: A music education perspective. International Journal of Music Education, 31(2), 202-216.

[16] Jay, G., & Jones, S. E. (2005). Whiteness studies and the multicultural literature classroom. Melus, 30(2), 99-121.

[17] Joseph, D., Nethsinghe, R., & Cabedo Mas, A. (2018). Creating multicultural music opportunities in teacher education: Sharing diversity through songs. Australian Journal of Teacher Education, 43(5), 32-47.

[18] Kaushik, V., Narayana, A., Satha, D. S., Grace, J., Elankavi, R., & Chavan, G. (2025). Generative AI for multicultural music education. ShodhKosh: Journal of Visual and Performing Arts, 6(5s).

[19] Krebs, K. C. (2020). Global, international, and intercultural learning in university classrooms across the disciplines. Research in Comparative and International Education, 15(1), 36-51.

[20] Lee, J. (2020). Promoting majority culture and excluding external ethnic influences: China's strategy for the UNESCO intangible cultural heritage list. Social Identities, 26(1), 61-76.

[21] Li, J. (2016). Application of national opera singing techniques on vocal music. In 2016 International Conference on Economy, Management and Education Technology (pp. 386-390). Atlantis Press.

[22] Li, Y. (2020). Study on the problems and countermeasures of Chinese multicultural music education. In International Conference on Arts, Humanity and Economics, Management (pp. 106-109). Atlantis Press.

[23] Liu, Q. (2022). Culturally and linguistically responsive teaching with Chinese characteristics? A multi-method study in southwest China [Doctoral thesis]. KU Leuven.

[24] Mellizo, J. M. (2019). Exploring the effect of music education on intercultural sensitivity in early adolescence: A mixed methods inquiry. Music Education Research, 21(5), 473-487.

[25] Parada-Cabaleiro, E., Batliner, A., & Schedl, M. (2022). An exploratory study on the acoustic musical properties to decrease self-perceived anxiety. International Journal of Environmental Research and Public Health, 19(2), 994.

[26] Parker, R. (2015). Multicultural music and learning. GEMS (Gender, Education, Music, and Society), 8(3), 20-27.

[27] Petrovic, J. E., & Caddell, A. (2020). Intensifying multicultural education through critical pedagogy, antiracism, and the need to unschool. In Oxford research Encyclopedia of education. Oxford University Press.

[28] Westerlund, H., Karlsen, S., & Partti, H. (2020). Visions for intercultural music teacher education. Springer Nature.

ISSN: 3079-6342

[29] Yap, J. H. (2025). The challenges and issues in artificial intelligence (AI) in vocal performance education: A comprehensive narrative review. International Journal of Research and Innovation in Social Science.

[30] Zhang, W. (2017). Multicultural ethnic music education in Communist China. International Journal of Multicultural Education, 19(3), 65-84.