Música Hodie | ISSN: 2317-6776 DOI: 10.5216/mh.v26.83504

Exploring the generative AI art technologies in vocal training for increasing aesthetic and emotional perceptions

Explorando as tecnologias de arte generativa da IA no treino vocal para aumentar as perceções estéticas e emocionais

Beibei Zhang
Harbin Normal University, Harbin, China
zhangbeibeihsd@edu.cn

Abstract: The objective of this article is to explore approaches to employing generative artificial intelligence and artistic technologies in vocal training to enhance aesthetic and emotional perception. Throughout the investigation, the potential utilization of various intelligent technologies for vocal education, including ChatGPT, LANDR, DeepSinger, and VOCALOID, was identified. It was determined that the VOCALOID (97.3) and DeepSinger (95.0) applications offer greater advantages for developing vocal skills. The research findings revealed that the application of DeepSinger enhanced emotional expressiveness in respondents' singing (attributed to its natural sound-processing capabilities). Additionally, VOCALOID contributed to shaping the aesthetics of singing by enabling voice generation and comparison of different performance variations. The study also identified challenges associated with the use of artificial intelligence technologies in education, including a lack of individualized consultations (32%), imprecise interpretation of musical compositions (28%), a deficit of creativity in teaching (23%), and the availability of moderately high-quality instructional materials (17%). At a higher level, the emotional and aesthetic perception of singing by audiences was achieved among respondents who utilized the DeepSinger

application during their training (average value: 0.83). Conversely, the least expressive singing was observed among respondents who utilized the ChatGPT application (average value: 0.5) in their training, primarily focused on developing theoretical rather than practical skills. The practical significance of this article lies in cultivating vocal singing skills through the utilization of artificial intelligence technologies such as ChatGPT, LANDR, DeepSinger, and VOCALOID.

Keywords: expressive singing. musical melismas. transformation of musical composition. virtual performers. vocal timbre.

Resumo: O objetivo deste artigo está associado à exploração de abordagens para o emprego de tecnologias artísticas de inteligência artificial generativa no treino vocal para melhorar a perceção estética e emocional. Ao longo da investigação foi identificado o potencial de utilização de diversas tecnologias inteligentes para a educação vocal, incluindo o ChatGPT, LANDR, DeepSinger e VOCALOID. Determinou-se que as aplicações VOCALOID (97,3) e DeepSinger (95,0) apresentam vantagens mais significativas no desenvolvimento das competências vocais. Os resultados da investigação revelaram que a aplicação do DeepSinger facilitou a melhoria da expressividade emocional no canto dos entrevistados (atribuído à sua capacidade de processamento de som natural). Além disso, o VOCALOID contribuiu para moldar a estética do canto, permitindo a geração de voz e a comparação de diferentes variações de performance. O estudo identificou ainda desafios associados à utilização de tecnologias de inteligência artificial na educação, incluindo a falta de consultas individualizadas (32%), a interpretação imprecisa de composições musicais (28%), o défice de criatividade no ensino (23%) e a disponibilidade de materiais instrucionais de qualidade moderadamente elevada (17%). A um nível mais elevado, a perceção emocional e estética do canto por parte do público foi alcançada entre os inquiridos que utilizaram a aplicação DeepSinger durante o treino (valor médio: 0,83). Por outro lado, o canto menos expressivo foi observado entre os

inquiridos que utilizaram a aplicação ChatGPT (valor médio: 0,5) na sua formação, com foco principalmente no desenvolvimento de competências teóricas, em vez de práticas. A importância prática deste artigo reside no desenvolvimento de competências de canto vocal através da utilização de tecnologias de inteligência artificial, como o ChatGPT, LANDR, DeepSinger e VOCALOID.

Palavras-chave: canto expressivo. melismas musicais. transformação da composição musical. intérpretes virtuais. timbre vocal.

Submitted on: July 27, 2025

Accepted on: November 18, 2025

Published on: February 2026

1. Introduction

Vocal singing is not only associated with the purity of execution but also with the conveyance of emotionality, which is formed through the precise perception of compositions. Expressiveness and immediacy are employed in vocal performance to engage listeners, contributing to a more expressive sound (Nikolai et al., 2023). Vocal singing should be linked to the development of creative abilities and analytical skills, fostering qualitative vocal performance. Achieving a high level of execution is possible through the use of artificial intelligence technologies, which can influence the development of vocal skills within more constrained time frames (Garcia, 2025).

Chinese vocal performance in the 21st century is characterized by a diversity of musical genres while simultaneously preserving the uniqueness of national traditions (Xiao et al., 2023). Significance is placed on the brightness of performance, the purity, and the emotionality of singing. Additionally, the naturalness of vocal singing is maintained and is associated with the preservation of unique elements of musical culture (Xu, 2022). The Chinese style of vocal performance involves the development of technological, physiological, and aesthetic aspects of singing, enabling the adaptation of high-quality vocal performance across various musical genres (Sáez-Velasco et al., 2024). Chinese singing is also characterized by a combination of Western vocal performance and national singing, expanding the boundaries of vocal performance. However, combining Western singing with Western singing is associated with significant differences in performance style, affecting rhythm, intonation, and overall artistic expression (Wu, 2025). Therefore, mindfulness should be incorporated into education to support comprehensive development.

The development of aesthetics and emotional singing depends on the style of execution and vocalization techniques. High-quality vocal performance is possible through the development of the natural voice, artistic perception, musical ear, breath control, and

rhythm (Temirbaev, 2025). A vocalist should possess a well-formed voice to ensure its flexibility. Emphasizing the expansion of vocal range allows for a more vibrant sound, enabling the interpretation of complex musical compositions. To facilitate quality education, artificial intelligence technologies can be utilized (Cui and Chen, 2023; Guo and Shiobara, 2022). These technologies enable teaching through audio and video lectures, influencing material visualization for better comprehension. The effectiveness of generative artificial intelligence technologies is associated with their ability to provide step-by-step training, thereby refining individual musical skills, including intonation (Cui and Chen, 2023).

Artificial intelligence technologies contribute to the development of diction, maintaining the correct tempo, shaping the expressive coloring of compositions, and enabling the conveyance of meaning. Artificial intelligence facilitates student interaction, as evidenced by feedback received during learning (Zhang, 2025). Feedback mechanisms may involve chatbots, virtual assistants, and voice bots, all powered by artificial intelligence. This approach enhances student motivation, thereby increasing the efficiency of the learning process (Nie and Ng, 2023). With the assistance of artificial intelligence, students can also develop creative skills, thereby eliminating constraints on task selection (Ji, 2022). Consequently, the integration of modern technologies into the educational process enhances the quality of the skills students acquire and fosters a more conscious perception of information.

1.1 Literature review

The development of emotional expressiveness in vocal singing must be associated with meticulous preparation, enabling orientation towards diverse mechanisms. Artificial intelligence technologies that facilitate neurovisualization offer approaches to achieving emotional singing. Contemporary technologies allow the exploration of vocal timbre and the classification of discernible emotions, intersecting voice interactions with music and shaping the processing of vocal data. It is also essential to consider vocal

interpretation methods in line with the selected performance technique. This will enable the creation of melodic nuances and contribute to achieving emotional resonance in performance (Wang et al., 2024). Intelligent technologies in vocal education ensure the transmission and encoding of sounds, as evidenced by the modeling of various data. Emotional singing can be achieved through an orientation toward multidimensional analysis, which contributes to accurate vocal expression. Focusing on more complex acoustic parameters aids in vocal training, subsequently enhancing its aesthetics. Artificial intelligence facilitates interaction with the human voice and enhances the possibilities of vocal performance. Moreover, it enables a more creative approach to vocal training (Creely et al., 2024).

The development of emotional vocal singing can be achieved through performing compositions from different musical genres. The educational approach facilitates the cultivation of students' musical literacy. An orientation towards folk, Western, Chinese opera, chamber, and pop music enables a more precise development of emotional expressiveness in performance. Various intelligent technologies contribute to proper vocal placement. Applications such as SwiftScales facilitate the development of melodic skills, Learn Singing enhances articulation precision, and Singscope promotes overall expressive vocalization (Li, 2024).

Intelligent technologies contribute to vocal education by leveraging peripheral algorithms to monitor vocal performance and influence the attainment of high-level vocal skills. These technologies are also associated with the assessment of vocal acoustics, impacting the formation of vocal speech. Students can track their performance quality, facilitating the development of both technical and emotional vocal skills (Mengchun and Zhixin, 2022).

Intelligent technologies facilitate a novel teaching style that influences communication and the cultivation of proper vocal skills. Adjustments to instruction can occur through feedback, impacting the qualitative execution of individual sounds. Additionally, artificial intelligence technologies enhance student motivation

and stimulate cognitive development. The acquisition of such skills enhances singing quality and encourages the exploration of individual performance strategies (Yu and Xiong, 2022).

Artificial intelligence technologies enable intelligent sound signal recognition, contributing to musical education. These technologies enhance musical notation throughout musical education, helping students memorize and recognize various rhythms. They also facilitate the synthesis of musical melodies, influencing the study of approaches to interpretation. In education, this approach can contribute to the identification of vocal segments and influence meaningful vocalization. The LSTM-based music learning model helps capture the distinctive features of musical performance, ensuring the accuracy and comprehensiveness of vocal ability assessment (Han, 2025). Computer technologies can be employed to simulate musical fragments, enabling personalized data in education and facilitating independent learning. Various parameters can be integrated into the learning process, such as musical sight-singing, evaluating the complexity of musical scores, and developing vocal performance sequences. This integration influences the development of vocal abilities, manifested in vocal technique and aesthetics (Zhe, 2021).

The cultivation of emotional expressiveness in singing is achievable through the meticulous examination of the musical text. Artificial intelligence technologies enable comparisons of emotion algorithms and the accuracy of their representations based on musical text. Implementing this approach involves studying aspects of noise, speed, and pitch in performance. Emotion classification during singing can also be achieved using audio files, thereby facilitating auditory perception of melodies. This ensures a comprehensive exploration of all musical aspects, contributing to the sequential development of vocal skills (Anchan et al., 2024).

The analysis of existing studies has revealed various approaches to utilizing artificial intelligence technologies to develop vocal skills. However, gaps in the published research are evident, particularly in the emphasis on sound alterations to

foster emotional and aesthetic aspects of singing. The specifics of how this can be implemented using contemporary technologies are addressed only superficially.

1.2 Problem statement

Vocal performance in China receives particular attention for its role in preserving national traditions that permeate across generations. The quality of vocal performance in China is intricately linked to technical aspects and emotional expression during execution. This association is rooted in the elevated artistic nature of musical compositions unique to Chinese culture. Novel approaches and technologies are employed in education to preserve this distinctive sound, influencing overall personal development and fostering an understanding of the nuances of traditional culture. This article aims to explore generative artificial intelligence artistic technologies in vocal education with the intent of enhancing aesthetic and emotional perception.

The research objectives were associated with the following points:

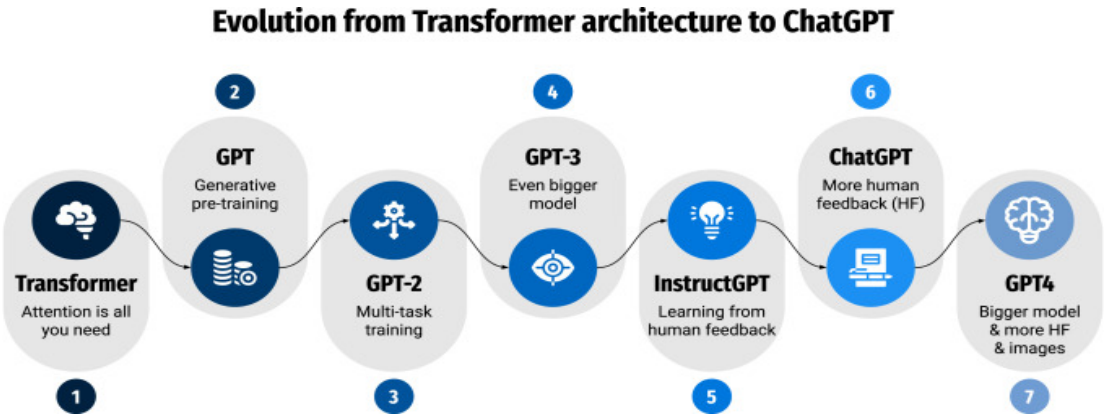
- To determine the functionality of the ChatGPT, LANDR, DeepSinger, and VOCALOID applications in facilitating the development of technical, emotional, and aesthetic skills in vocal singing;
- To assess the levels of emotional and aesthetic development in vocal singing among respondents, utilizing different artificial intelligence technologies;
- To identify challenges encountered by respondents during vocal training;
- To discern respondents' emotional and aesthetic perceptions of vocal performances across those who used different artificial intelligence technologies.

2. Methodology

2.1 Research design

The first stage of the research involved an examination of various aspects of generative artificial intelligence art technologies to explore their potential applications in vocal performance education. From a multitude of technologies, the selection was narrowed to ChatGPT, LANDR, DeepSinger, and VOCALOID, as they do not belong to the same category and do not share similar functionalities. During the selection of artificial intelligence technologies, considerations were made based on their popularity and broad applicability. The potential for developing vocal skills through an understanding of the theoretical or practical aspects of singing was also taken into account. The authors ensured that the chosen artificial intelligence technologies were accessible for use. These four applications were selected from a pool of 70 different technologies. Some applications were excluded based on an assessment of their capabilities, which precluded the development of singing skills during actual vocalization. Other applications were excluded after practical implementation because they hindered comprehensive vocal skill development. Developing the aesthetics of vocal singing proves more challenging with ChatGPT, given the lack of high-quality audio for understanding the harmonious combination of musical elements. ChatGPT serves more as an evaluative tool for understanding vocal music theory than a practical instrument for training. This is because the selection of timbres and rhythms for harmonious performance is directly guided by the vocalist (Figure 1).

Figure 1 - The functionality of the ChatGPT Application

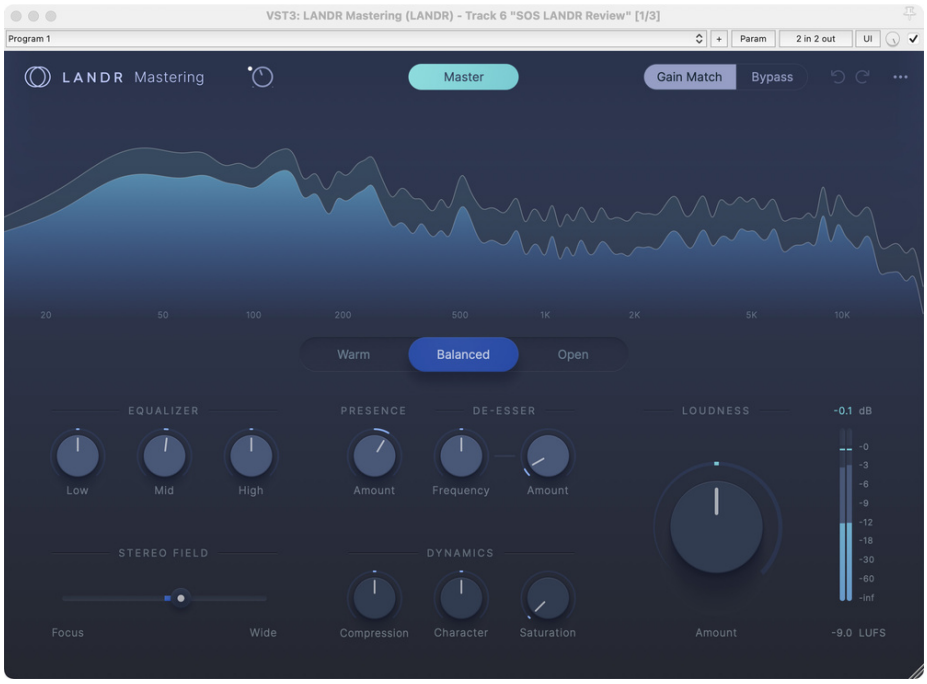


The adaptation of ChatGPT technology for vocal training involved respondents completing assessments to determine their vocal abilities and receiving a structured plan for developing technique, aesthetics, and emotional expression in performance. ChatGPT technology enabled the provision of tailored recommendations for improving singing, including guidance on selecting practical exercises.

The LANDR technology can help create an accurate representation of the musical composition's performance. This is attributed to its ability to ensure proper elaboration of the musical composition by the vocalist. The artificial intelligence technology in LANDR adapts a specific composition to the vocalist's capabilities, taking into account various sound effects. The automated search for melismas in a musical piece contributes to the development of approaches to emotional performance. The use of LANDR technology facilitates ear training by adapting the melody to the vocalist's voice. The application is most significant for enhancing the emotional aspects of vocal performance, as it establishes an emotional connection with the music through individual tools. Achieving technical precision in performance is possible by modifying the composition's sound characteristics and by considering proper dynamics and tonality in execution. High-quality sound can be achieved by analyzing all musical nuances of the composition with LANDR. However, for aesthetic

development, students should rely on their abilities, fostering aesthetic performance (Figure 2).

Figure 2 - Features of the LANDR Technology Operation

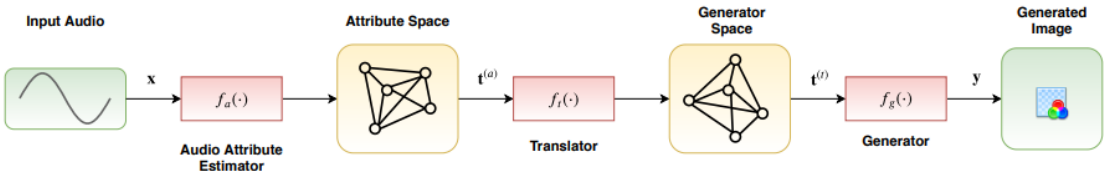


The LANDR technology was adapted for training purposes by initially processing vocal performances from a group of vocalists and analyzing intonation. This approach enabled students to compare their performances with original recordings and to select automated methods for enhancing sound quality, including the incorporation of additional effects.

The artificial intelligence technology DeepSinger can facilitate vocal training, albeit not in real time, by processing individual vocal performance recordings. Developing vocal skills is possible by capturing the performer's timbre and tailoring training to their vocal capabilities. Enhancing vocal technique and expressiveness can be achieved through integrated tools that monitor the accuracy of musical composition performance by controlling pitch and duration. DeepSinger enables the development of more advanced vocal skills by complicating rhythms or specific compositional

segments. Through accurate compositional analysis, issues with sound-text harmony affecting vocal harmony can be identified and eliminated. The transformation of compositions enables their adaptation to the performer's vocal abilities, mitigating complex performance aspects. This approach influences vocal performance quality by emphasizing sound patterns and the ability to use Chinese, Cantonese, and English compositions. Attaining an average performance rating motivates vocalists to achieve higher results (Figure 3).

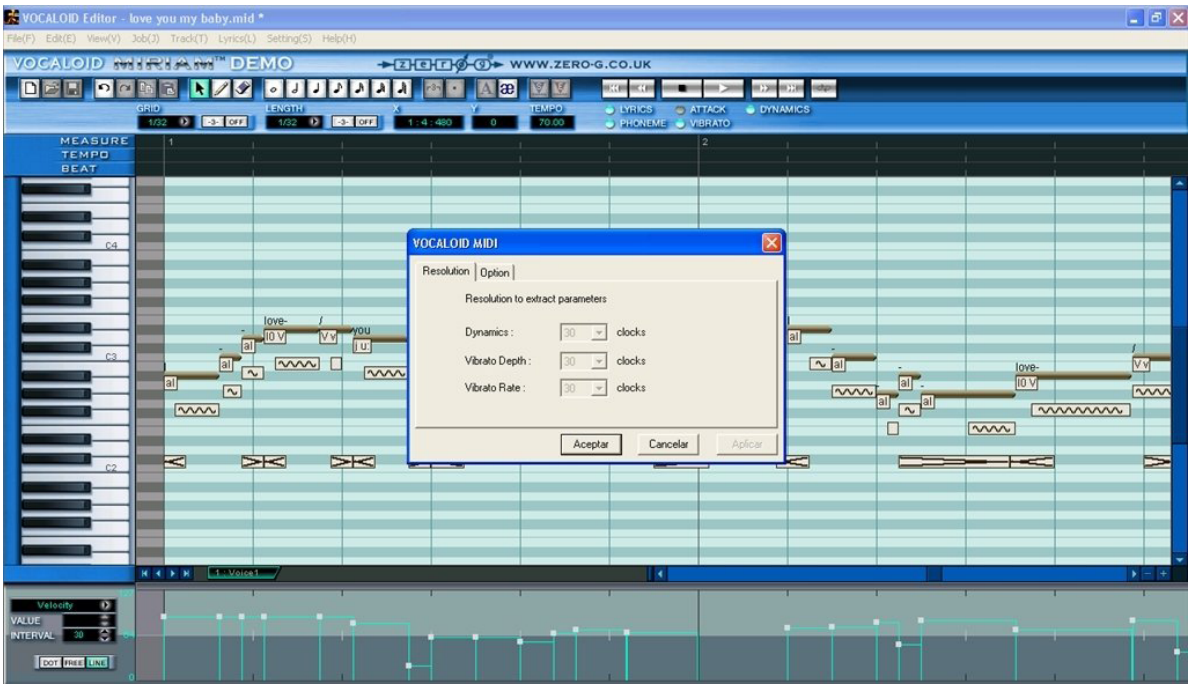
Figure 3 - The functionality of the DeepSinger Artificial Intelligence Technology



The adaptation of the DeepSinger application for vocal training involved creating vocal performance recordings and tailoring the learning process to respondents' proficiency levels.

The VOCALOID technology enhances vocal performance by thoroughly analyzing the voice. During training, it is possible to develop the correct singing characteristics, which involve analyzing timbre and emotionality. This approach allows the rendition of diverse compositions, taking into account their unique features. The application also facilitates the representation of virtual performers' vocals. With such functionality, it is possible to focus on the nuances of compositions performed by other vocalists, promoting improvement in vocal expressiveness and the alteration of vocal timbres. The technique of execution can be ensured through full speech synthesis, which shapes natural singing. Synthesizing the performed voice enables the composition to be divided into smaller fragments, improving handling. Processing musical segments from one's own performance and from professional performers' helps establish similarities and differences, ensuring the quality of vocal interpretation (Figure 4).

Figure 4 - VOCALOID Software Functionality



The synthesis of respondents' performing voices enabled the adaptation of VOCALOID technology for vocal training. The detailed distribution of vocal parameters influenced the selection of an appropriate training program.

After selecting the applications, their quality in terms of vocal technique, emotional expression, and aesthetics was considered, focusing on their potential. The development of vocal technique encompassed voice placement and the accurate rendition of musical composition fragments. Emotional expression in singing was associated with conveying artistic fragments, lyrical digressions, and musical elements. The performance's aesthetics implied the harmonious execution of melodies, combining vocal technique and emotional expressiveness. The determination of technique, emotional expression, and aesthetics of reproduction was defined through the practical application of technologies, enabling the identification of the most pronounced functionalities in each.

The study's second phase involved conducting vocal training for participants using various artificial intelligence technologies.

The training spanned 4 months, aiming to cultivate technique, emotional expression, and overall aesthetics in the respondents' vocal performances. Throughout the training period, participants focused on the functionalities of one of the applications (ChatGPT, LANDR, DeepSinger, or VOCALOID) and engaged in exercises designed to facilitate learning. The training was supervised by instructors, ensuring coordinated educational processes. After the 4-month training period, participants were assessed for their level of development in emotionality and aesthetics. The results were collected continuously throughout the training, eliminating reliance on outcomes alone. Pedagogues assigned the assessments to determine the emotional and aesthetic skills developed during the training. The participants' abilities were graded, further reflected in additional Wilks' coefficient calculations (Chen, 2022). This approach facilitated an additional statistical comparison of the obtained results.

The third stage of the research aimed to identify the challenges students encounter during the learning process. These challenges were associated with the functionalities of the applications used in education. A list of issues was directly provided by respondents, with detailed justifications, using a survey method. The survey was conducted over one day among respondents via WeChat. The article's authors (based on educators' opinions) presented potential problems that the respondents might have faced during the learning process. Respondents could choose from the provided list of issues or express their own views on educational challenges. The reliability of the results was assessed using data collected from instructors two weeks before and after the completion of training. This approach enabled a more comprehensive evaluation of the outcomes and facilitated the identification of previously unaccounted gaps.

The fourth stage of the research involved determining the quality of song performances based on audience opinions. The audience was required to listen to students from different groups as they sang to assess their quality, harmony, and overall

aesthetic performance. The listening sessions for various groups of respondents were conducted by the same audience over 4 days to eliminate variations in opinions.

The perception coefficient was calculated based on the results obtained using the methodology of Livingstone et al. (2014) for assessing aesthetic perception and the method of Bachorowski and Owren (1995) for evaluating emotional perception, as applied by the viewers for their assessments. The Livingstone et al. (2014) method for assessing aesthetic perception was linked to the determination of the artistry of performance. The process involved an initial introduction to the musical works being performed, which allowed viewers to analyze the original performances of renowned vocalists in recordings, study performance style, and observe stage behavior. Original performances were compared to those of the respondents, with attention given to their emotional expression and the balance of their performances. Additionally, the evaluation focused on the freedom of improvised performance and concentration during execution.

The Bachorowski and Owren (1995) method for evaluating emotional perception assessed viewers' emotions before and after listening to vocal performances. This approach aimed to assess statistical differences between viewers' baseline reactions and their responses after the performance, which involved the respondents completing vocal tasks. To assess the reliability of results obtained from viewers, the interactive program Affectiva was used, which enabled automatic tracking of viewers' emotions. This enabled an analysis of the emotionality of vocal performances, based on both viewers' responses and data from the Affectiva program.

2.2 Sample

During the research, 185 respondents, who had been studying vocal singing for 2 years, were recruited. The selection of respondents involved choosing vocalists with traditional and classical music backgrounds, ensuring equal initial conditions for all participants in the study. Respondents were selected based on

applications submitted through the Weibo platform in response to the provided advertisement. The 2-year singing education experience had to be connected with professional training (music school, extracurricular classes, etc.), excluding self-learning. The respondents' ages ranged from 15 to 17 years.

Additionally, 185 respondents were divided into four groups based on the technologies used in vocal training. Group 1 consisted of 47 participants who used ChatGPT technology during their training. Group 2 included 46 participants who trained with the LANDR technology; Group 3 included 47 participants who trained with the DeepSinger technology; and Group 4 included 46 participants who trained with the VOCALOID technology. The distribution of respondents into groups was not dependent on a specific level of knowledge within any given group. In all groups, respondents were evenly distributed by vocal skill. Each group had approximately 65% of participants with advanced vocal skills and 45% with above-average vocal skills. This distribution eliminated the random influence factor on the final results and was aligned with the use of a targeted respondent group. Respondents' abilities were assessed through listening sessions. The listening process involved performing selected songs, which were of equal difficulty. The selection of trained vocalists allowed for a focus on the primary research task and the evaluation of artificial intelligence technologies for vocal training. The combination of respondents with high and intermediate vocal performance levels enabled the assessment of the effectiveness of the technologies across different skill levels.

Additionally, 357 spectators were involved in the study and were selected via the Weibo social network. The selection of spectators was limited solely by age criteria (21–49 years), contributing to a conscious perception of the musical compositions. Spectators aged 49 or older were not included in the study, as their applications were excluded from potential participants.

2.3 Statistical processing

Significant emphasis was placed on calculations in the research, thereby enhancing the precision at each stage of the study. In the course of the calculations, particular attention was given to initially categorizing respondents' responses, which directly influenced the computations. The processing of the final results was facilitated using the TableEdit program.

2.4 Ethical issues

Ethical considerations involved the use of accurate information that corresponded to the research theme and was obtained solely within the framework of the study. The authors excluded the borrowing or falsification of others' results that were unrelated to the research (Committee on Publication Ethics, 2021).

2.5 Research limitations

The study's limitations include examining only technologies and artificial intelligence for vocal training and excluding traditional teaching methods. Future research endeavors are planned to explore diverse approaches to vocal singing to determine their effectiveness.

3. Results

The variety of generative artificial intelligence art technologies impacts the facilitation of the learning process, whether through self-guided learning or group instruction. To analyze the potential use of these art technologies in vocal singing, an initial examination was conducted. During the analysis, emphasis was placed on the potential for developing vocal performance techniques and aesthetics (Table 1).

Table 1 - Comparison of the Specifics of Artificial Intelligence Technologies for Vocal Training

Name of Artificial Intelligence Technology	Technique	Emotionality	Aesthetics	Overall Score (Averaged Metrics)
ChatGPT	86.3	84.5	83.7	84.8
LANDR	90.2	96.3	83.7	90.1
DeepSinger	97.9	92.7	94.5	95.0
VOCALOID	95.2	93.1	92.8	97.3

ChatGPT technology contributes to deep learning in music, yet its functionality is predominantly oriented toward theoretical aspects. This is attributed to the absence of qualitative features for reproducing digital audio, which affects the quality of the music played. The sound is more monophonic, limiting its ability to convey musical aspects. However, ChatGPT enables the development of precise technical aspects during vocal training. The analysis of musical compositions using artificial intelligence helps develop an understanding of their vocal complexity or simplicity. Text prompts facilitate the analysis of composition, enabling the study of artistic embellishments that influence the overall emotional expression in singing. It also contributes to comprehending the technical accuracy of the performance.

Based on the functionality of musical compositions, their practical impact was determined to assess overall aesthetics and emotional performance. The results were obtained through student training using various technologies, including ChatGPT, LANDR, DeepSinger, and VOCALOID (Table 2).

Table 2 - Determination of Acquired Skills in Emotional and Aesthetic Singing among Students of Different Groups

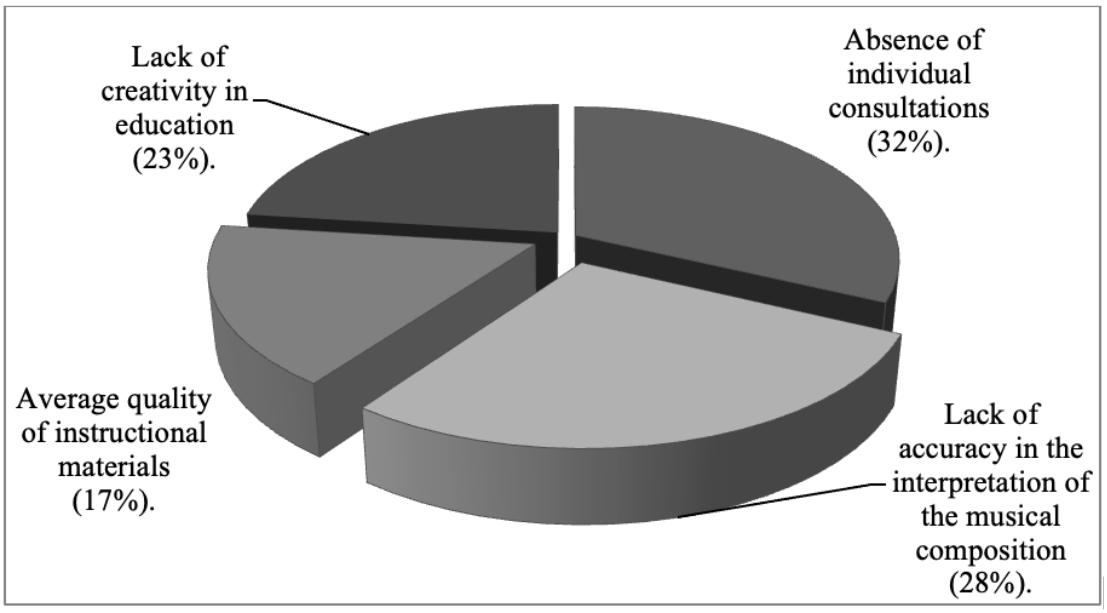
Groups of Students Depending on the Used Technologies	Emotional Singing		Aesthetic Singing		F
	M	DT	M	DT	
ChatGPT	15.3	6.7	15.9	6.4	0.72
LANDR	17.6	6.1	16.1	5.7	0.77
DeepSinger	18.3	5.63	17.8	5.8	0.83
VOCALOID	17.4	5.9	18.1	5.54	0.81

After training the respondents in vocal performance, it was observed that the generative artificial intelligence technologies used had varying effects on singing quality. Notably, the DeepSinger technology primarily contributed to the development of emotional aspects in vocal performance among the respondents. This is attributed to its capability to generate compositions for execution through natural sound processing. Focusing on the characteristics of the transformed musical composition enables a more qualitative singing experience that aligns with the vocal capabilities of the individuals. The VOCALOID application facilitated the attainment of the highest aesthetic quality in vocal singing. The aesthetics of vocal singing were associated with the combination of artistic and technical aspects of vocal performance. Achieving this was possible through the utilization of vocal imitation from other performers. By generating vocal recordings, it was feasible to compare different performance variations and identify deviations in the respondents' vocal singing. Additionally, students who used VOCALOID technology in their training could achieve greater precision in selecting artistic embellishments, thereby influencing the overall aesthetic of the performance.

Implementing the LANDR application has facilitated achieving average indicators of emotional expression and overall aesthetic performance. This capability is associated with the ability to ensure precise preparation of a musical composition for execution; however, it is impacted the absence of approaches to training in vocal technique. Notwithstanding this, appropriate preparation of a musical composition enables the comprehension of the specificity of its interpretation. The artificial intelligence ChatGPT has played a minimal role in developing vocal skills. The outcomes are linked to ChatGPT's ability to address theoretical aspects of learning exclusively, requiring an understanding of the technical facets of singing and their practical implementation.

Subsequently, the study sought to identify the challenges students encounter in learning to sing through the use of artistic technologies based on generative artificial intelligence (Figure 5).

Figure 5 - Challenges Encountered by Respondents in the Process of Vocal Training, Considering the Specifics of Artificial Intelligence Technologies



The analysis of respondents' opinions revealed that the most prevalent issues were related to the use of intelligent technologies, specifically ChatGPT and LANDR. While these technologies lack specialized functionality for vocal training, they may contribute to musical education. The primary challenge stems from the absence of individual consultations, which negatively impacts understanding of the most appropriate methods for interpreting melodies. Additionally, instructional difficulties were linked to insufficient precision in interpreting musical compositions. These issues manifested as the imprecise use of artistic embellishments intended for expressive vocal performance. Furthermore, the presented problem could manifest as compositions being transformed for execution with increased complexity, which does not align with vocalists' acquired skills.

A common problem was the lack of creativity in education. The process involved examining a single interpretation of a composition, neglecting alternative perspectives, which could affect the development of appropriate approaches to interpreting it. The moderate quality of some instructional materials did

not provide a clear understanding of the technical or practical aspects of reproducing melodies, necessitating the search for supplementary materials to improve vocal performance.

Following the instructions, the audience's level of perception of vocal performances was assessed. Results were obtained from various student groups engaged in vocal training (Table 3).

Table 3 - Audience Perception Level of Aesthetics and Emotionality in Vocal Singing by Students Depending on the Utilized Artificial Intelligence Technologies in Education

Student groups based on the technologies used	Emotional perception			Aesthetic perception			F
	The value of the perception coefficient.	M	DT	The value of the perception coefficient	M	DT	
ChatGPT	0.57	14.8	6.9	0.43	14.1	7.4	0.71
LANDR	0.74	15.6	6.3	0.76	15.8	6.2	0.84
DeepSinger	0.81	17.9	5.6	0.85	18.3	5.1	0.89
VOCALOID	0.74	15.5	6.2	0.78	15.9	6.0	0.87

The analysis of audience evaluations of vocal performance quality revealed differences among distinct groups of vocalists. Performers who used the DeepSinger application during training achieved a higher quality of execution, as reflected in audience ratings. The performances of this group of vocalists were characterized by greater expressiveness, emotional depth, and a capacity to convey musical imagery. Following closely in terms of audience perception of emotional singing were students who received training through the VOCALOID and LANDR applications. However, their performances lacked a cohesive aesthetic. Students trained with ChatGPT, to a lesser extent, engaged the audience. Their singing lacked expressiveness and emotional resonance. Respondents were more focused on performance technique than adhering to a unified aesthetic in their singing.

4. Discussion

Artificial intelligence technologies have wide applications in the contemporary education system. They offer an alternative to traditional vocal training, focusing on voice placement, arrangement use, and bel canto technique. Internet technologies enable a networked approach to vocal training, characterized by the sequential development of vocal skills. Training incorporates audio and video methods, allowing a more thoughtful educational approach grounded in specific examples (Xu and Xia, 2023). Vocal training can be achieved through intelligent analysis of musical composition structures and reproducible sounds. Analysis is essential for understanding the specificity of vocal notes and for facilitating real-time tracking of their accuracy in execution. Moreover, accuracy in score tracking is reflected in determining the quality of vocal singing. Students contribute to a better understanding of vocal singing nuances, leading to the motivated development of vocal skills (Ding and Zhang, 2024). The analysis of published works is based on the study of artificial intelligence's capabilities for learning musical nuances and the availability of accurate examples for guidance in training. In our article, attention was focused on enhancing the aesthetic and emotional perception of musical compositions.

The musical education process should not only focus on voice placement but also on the emotive aspects of singing. Modernizing the education system through contemporary technologies enables the development of musical knowledge in shorter durations. Achieving a beautiful sound is possible through breath control techniques. The use of digital technologies in education is associated with body coordination, breath training, and coordination of the vocal apparatus. Training has led to high-quality vocal performances of folk and pop compositions (Wang, 2024). In the analyzed literature, the development of vocal singing skills, including aesthetic aspects of performance, is linked to the specifics of the vocal voice. In our article, we emphasized studying

the characteristics of ChatGPT, LANDR, DeepSinger, and VOCALOID for the development of techniques, emotional expression, and aesthetic aspects of vocal singing.

Artificial intelligence technologies contribute to a more targeted approach to education, reflected in higher performance. Modernizing the music education system can be associated with the development of auditory skills, enhancing the perception of the nuances of playing music by ear. It also facilitates real-time assessment of singing, reflected in listening and the identification of chords, intervals, and musical genres. The development of vocal skills is achievable through a structured learning process and continuous support, providing personalized education. This approach can be implemented using technologies such as VoCo Vocal Coach, Swiftscales Vocal Trainer, and VocaLive CS (Shi, 2023). The use of mobile technologies and platforms has a positive impact on student motivation. Comparisons between traditional teaching and teaching with artificial intelligence (AI) technologies have shown clear advantages for the latter. The use of applications such as Yousician and Voice Training: Learn To Sing has influenced the development of vocal skills and clarity in vocal singing. It has been determined that student motivation plays a crucial role in the development of singing skills (Wang, 2022). In contrast to the presented works, our article focuses on the analysis of vocalists' voices rather than musical structures. This allows the learning process to adapt to the level of the vocalists.

A promising direction in music education is the use of computer technologies in vocal training. The Melodyne sound editor can be employed to develop technical vocal skills and serves as a comprehensive educational resource. The Melodyne application allows for the diagnosis of singing quality and corrects vocal performance. Vocal pitch recognition mechanisms contribute to its editing, influencing the intensity of developing proper vocal skills. The Melodyne application enables adjustments to pitch accuracy and rhythmic execution. The enhancement of these skills can be achieved through the creation of musical associations (Villanueva et

al., 2024). The effectiveness of artificial intelligence technologies has been demonstrated in the analyzed works through a comparison with traditional education. In our article, specific technologies for teaching (ChatGPT, LANDR, DeepSinger, VOCALOID) were selected, and their significance for the development of aesthetics and performance ethics was determined based on responses from respondents and the audience's perception of singing.

During vocal training utilizing digital technologies, it is imperative to ensure the acquisition of digital skills. Interactive technologies in education facilitate visual and convenient learning. A comparison between traditional and interactive learning has revealed distinctions between the approaches. The effectiveness of interactive learning was 10.34% higher, attributed to systematicity, professionalism, and modernity. The process has enabled enhanced learning capabilities that can be integrated into the educational system (Tang, 2024). Information technologies in musical education allow the analysis of students' developed abilities by integrating musical subject knowledge with an informational platform. It is possible to analyze vocal signals for singing purity throughout the training. The conduct of musical experiments is manifested in the development of students' vocal skills (Li et al., 2025). In our work, there is no comparison between traditional and interactive learning; however, we compare the possibilities for vocal development through training with different technologies (ChatGPT, LANDR, DeepSinger, VOCALOID).

The analysis of scholarly articles has revealed that these studies explore the intricacies of musical education employing intelligent technologies. However, their advantages are delineated based on a comparative assessment with the traditional educational system. In our article, we examined artificial intelligence technologies and proposed their integration into vocal training to enhance vocal aesthetics and emotional expression. Technologies such as ChatGPT, LANDR, DeepSinger, and VOCALOID were selected for investigation. Throughout the instructional process, their theoretical and practical significance for the development of vocal

skills was established. Moreover, the educational framework incorporated the assessment of respondents' perceptions of vocal aesthetics and of the audience's emotional expression.

5. Conclusions

The research was conducted with the defined objectives that aligned with the research goal. The study examined the advantages of artificial intelligence technologies in developing vocal techniques, emotional expression, and aesthetics. The applications selected for analysis were ChatGPT, LANDR, DeepSinger, and VOCALOID, allowing for the identification of their differences. It was established that applications VOCALOID (=97.3) and DeepSinger (=95.0) exhibited more suitable functionality for the development of vocal singing skills. The high efficacy of artificial intelligence technologies is attributed to their ability to control pitch, specific sound characteristics, and more. The LANDR application (LANDR (=90.1)) is more closely associated with the development of the capability to alter musical compositions to match vocalists' voices. Still, it does not affect the transformation of vocal singing. The ChatGPT application (=84.8) facilitates an understanding of the theoretical aspects of vocal singing but does not fully enable the development of practical skills.

After completing the training, vocalists who used the DeepSinger application (F=0.83) developed emotional expression, and those who used VOCALOID (F=0.81) exhibited enhanced vocal aesthetics. The respondents' results were linked to their ability to compare their vocal singing with the exemplary performances of renowned vocalists.

The research also involved identifying the challenges students encountered during the learning process. The study yielded average results among respondents who underwent vocal training using various artificial intelligence technologies (ChatGPT, LANDR, DeepSinger, VOCALOID). The results indicated that issues were associated with the absence of individual consultations, a lack

of precision in interpreting musical compositions, a deficiency in creativity in education, and a moderate quality of instructional materials. The audience's perception of the respondents' singing demonstrated an advantage in aesthetics and expressiveness among vocalists who utilized the DeepSinger application in their training.

The practical significance of this work lies in expanding the possibilities of vocal education through the use of artificial intelligence art technologies. Research prospects may be connected to discovering the potential applications of these technologies for the instruction of elementary school children and university students.

References

ANCHAN, A.; MANASA, G. R.; Pinto, J. P. Gender based real time vocal emotion detection. **International Journal of Intelligent Systems and Applications in Engineering**, v. 12, n. 3s, p. 282–289, 2024. https://www.ijisae.org/index.php/IJISAE/article/view/3706

BACHOROWSKI, J.-A.; OWREN, M. J. Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. **Psychological Science**, v. 6, n. 4, p. 219–224, 1995. https://doi.org/10.1111/j.1467-9280.1995.tb00596.x

CHEN, W. Design of music teaching system based on artificial intelligence. **Mathematical Problems in Engineering**, v. 2022, p. 2627395, 2022. https://doi.org/10.1155/2022/2627395

Committee on Publication Ethics. **Our Organisation**, 2021. https://publicationethics.org/about/our-organisation

CREELY, E.; HENRIKSEN, D.; HENDERSON, M.; MISHRA, P. **The Staging of AI: Exploring Perspectives about Generative AI, Creativity, and Education**. Rochester, NY: SSRN, 2024. https://doi.org/10.2139/ssrn.5008581

CUI, Z.; CHEN, C.-K. Art and history go hand in hand: The evolution of Chinese national vocal music from the Yan'an period to reform and opening up. **Heranca – History, Heritage and Culture Journal**, v. 6, n. 1, p. 61–72, 2023. https://doi.org/10.52152/heranca.v6i1.784

DING, Z.; ZHANG, B. Leveraging intelligent big data technology for optimizing vocal singing training based on embedded systems framework. **Computer-Aided Design and Applications**, v. 21, n. S8, p. 196–211, 2024. https://doi.org/10.14733/cadaps.2024.S8.196-211

GARCIA, M. B. The paradox of artificial creativity: Challenges and opportunities of generative AI artistry. **Creativity Research Journal**, v. 37, n. 4, p. 755-768, 2025. https://doi.org/10.1080/10400419.2024.2354622

GUO, K.; SHIOBARA, M. Natural, tender and unstrained: Teresa Teng in Chinese and Japanese music education. **Journal of Popular Music Education**, v. 6, n. 2, p. 193–205, 2022. https://doi.org/10.1386/jpme_00084_1

HAN, Y. Exploring a digital music teaching model integrated with recurrent neural networks under artificial intelligence. **Scientific Reports**, v. 15, Art. no. 7495, 2025. https://doi.org/10.1038/s41598-025-92327-8

JI, K. Classification method of emotional characteristics of vocal singing under multicultural values. **Journal of Sensors**, v. 2022, Art. no. 8168970, 2022. https://doi.org/10.1155/2022/8168970

LI, R.; HOANG, T.; TRIEU, V.; KHALAJZADEH, H. Integrating immersive technology in music education for coping with music performance anxiety. In: **Proceedings of the Twenty-Third Annual Pre-ICIS Workshop on HCI Research in MIS**. Bangkok: AIS, 2025, p. 1–6. https://aisel.aisnet.org/sighci2024/16

LI, Y. How does the music that students listen to affect their level of musical literacy? Comparative analysis of Chinese students' musical literacy formed with modern online technologies in the context of music

and non-music academic majors. **Interactive Learning Environments**, v. 32, n. 8, p. 4688–4702, 2024. https://doi.org/10.1080/10494820.2023.2205895

LIVINGSTONE, S. R.; CHOI, D. H.; RUSSO, F. A. The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. **Frontiers in Psychology**, v. 5, Art. no. 156, 2014. https://doi.org/10.3389/fpsyg.2014.00156

MENGCHUN, Y.; ZHIXIN, X. Liubai technique in the real-time electroacoustic Chinese art song 'Lang Tao Sha'. **Organised Sound**, v. 27, n. 3, p. 338–345, 2022. https://doi.org/10.1017/S1355771822000486

NIE, W.; NG, W. Analysis of multimedia feature extraction technology in college vocal performance teaching mode based on multimodal multimedia information. **International Journal of Web-Based Learning and Teaching Technologies**, v. 18, n. 2, p. 1–12, 2023. https://doi.org/10.4018/IJWLTT.329604

NIKOLAI, H.; KISIEL, M.; PENG, Y. Vocal education in the context of intercultural communication: Experience of teaching Chinese students. **International Journal of Chinese Education**, v. 12, n. 1, p. 1–12, 2023. https://doi.org/10.1177/2212585X231169741.

SÁEZ-VELASCO, S.; ALAGUERO-RODRÍGUEZ, M.; DELGADO-BENITO, V.; RODRÍGUEZ-CANO, S. Analysing the impact of generative ai in arts education: A cross-disciplinary perspective of educators and students in higher education. **Informatics**, v. 11, n. 2, p. 37, 2024. https://doi.org/10.3390/informatics11020037

SHI, Y. The use of mobile internet platforms and applications in vocal training: Synergy of technological and pedagogical solutions. **Interactive Learning Environments**, v. 31, n. 6, p. 3780–3791, 2023. https://doi.org/10.1080/10494820.2021.1943456

TANG, C. The construction and use of a higher education singing art teaching system combining digital multimedia art and information processing technology. **Computer-Aided Design and Applications**, v. 21, n. S2, p. 39–43, 2024. https://doi.org/10.14733/cadaps.2024.S2.39-53

TEMIRBAEV, B. Singing performance arts formation: The significance of education and the creative development of youth. **Jurnal Teknologi Pendidikan**, v. 2, n. 3, p. 1–6, 2025. https://doi.org/10.47134/jtp.v2i3.1434

VILLANUEVA, J.; ILARI, B.; HABIBI, A. Long-term music instruction is partially associated with the development of socioemotional skills. **PLoS ONE**, v. 19, n. 7, Art. no. e0307373, 2024. https://doi.org/10.1371/journal.pone.0307373

WANG, H.; JAMALI, S. S. B.; MUHID, S. H. B. Fusion and innovation: Exploring the application and effects of the Italian Bel Canto technique in enhancing the expressiveness of Chinese national vocal performance. **Multidisciplinary Reviews**, v. 8, n. 2, Art. no. e2025065, 2024. https://doi.org/10.31893/multirev.2025065

WANG, Y. Music education: Which is more effective – Traditional learning or the introduction of modern technologies to increase student motivation? **Learning and Motivation**, v. 77, Art. no. 101783, 2022. https://doi.org/10.1016/j.lmot.2022.101783

WANG, Y. The effectiveness of innovative technologies to manage vocal training: The knowledge of breathing physiology and conscious control in singing. **Education and Information Technologies**, v. 29, p. 7303–7319, 2024. https://doi.org/10.1007/s10639-023-12108-6

WU, H. Modern Chinese and European vocalist training methods: An exploration of the involvement of bachelors with the help of IT' sources. **European Journal of Education**, v. 60, n. 1, Art. no. e12847, 2025. https://doi.org/10.1111/ejed.12847

XIAO, M.; AMZAH, F.; RONG, W. Experience of beauty: Valuing emotional engagement and collaboration in teacher-child storytelling activities. **International Journal of Learning, Teaching and Educational Research**, v. 22, n. 2, p. 165–187, 2023. https://www.ijlter.net/index.php/ijlter/article/view/1524

XU, F.; XIA, Y. Development of speech recognition system for remote vocal music teaching based on Markov model. **Soft Computing**, v. 27, n. 14, p. 10237–10248, 2023. https://doi.org/10.1007/s00500-023-08277-8

XU, J. The most influential elements of the Chinese traditional music culture before the Ming Dynasty in Southeast Asia. **Critical Arts**, v. 36, n. 5-6, p. 177–190, 2022. https://doi.org/10.1080/02560046.2023.2180046

YU, P.-J.; XIONG, M.-Z. Remote vocal singing course design based on embedded system and internet of things. **Mobile Information Systems**, v. 2022, Art. no. 8712081, 2022. https://doi.org/10.1155/2022/8712081

ZHANG, Y. Increasing emotional perception in academic singing during vocal performance: The use of AI solutions. **International Journal of Human-Computer Interaction**, vol. 41, n. 19, p. 12086-12094, 2025. https://doi.org/10.1080/10447318.2025.2452213

ZHE, T. Research on the model of music sight-singing guidance system based on artificial intelligence. **Complexity**, v. 2021, Art. no. 3332216, 2021. https://doi.org/10.1155/2021/3332216.

Research ethics committee approval

The study protocol was approved by the Ethics Committee of Harbin Normal University (protocol No. 178 of Sept 5, 2023). The study was conducted in accordance with the rules of the Declaration of Helsinki. All subjects gave written informed consent before participation.

Publisher

Federal University of Goiás. School of Music and Performing Arts. Graduate Program in Music. Publication in the Portal of Periodicals UFG.