Computer Fraud and Security
ISSN (online): 1873-7056

Design and Implementation of Vocal Music Creation and Arrangement System Aided by Artificial Intelligence

Shiming Zhou
School of Music and Media, Anyang University,
No. 599, Zhonghua Road, Wenfeng District, Anyang City, Henan Province, 455000.
*Corresponding author:Shiming Zhou
Email address: ganbianyundou886@163.com

Abstract

With the rapid advancement of artificial intelligence (AI), its application in music creation and arrangement has gained significant attention. This paper presents the design and implementation of a vocal music creation and arrangement system aided by AI. The system integrates deep learning models, natural language processing, and digital signal processing techniques to generate and arrange vocal melodies based on user input. It employs recurrent neural networks (RNNs) and transformer-based architectures to analyze musical patterns, harmonization, and vocal synthesis, ensuring a seamless and coherent musical composition.

The system features an intuitive user interface, enabling composers, musicians, and amateurs to generate vocal melodies with minimal manual intervention. Additionally, it supports adaptive style transfer, allowing users to customize the genre and emotional tone of their compositions. The AI-driven arrangement module enhances the generated melodies by automatically structuring harmonies and accompaniments based on predefined musical styles.

Experimental results demonstrate the system's effectiveness in producing high-quality vocal arrangements with realistic synthesis and stylistic versatility. The proposed approach significantly reduces the time and expertise required for music composition, making AI-assisted vocal music creation more accessible and efficient. This study provides valuable insights into the fusion of AI and music composition, highlighting future possibilities for AI-driven artistic innovation.

Keywords: Artificial intelligence, vocal music, music arrangement, deep learning, AI composition, music synthesis.

Introduction

Music composition and arrangement have traditionally required extensive expertise and creative intuition. However, with advancements in artificial intelligence (AI), computational models have been increasingly employed to assist musicians in generating and arranging music. AI-driven music composition tools have the potential to revolutionize the creative process by automating certain aspects of melody generation,

harmonization, and vocal synthesis, making music production more accessible to both professionals and amateurs.

In recent years, deep learning techniques, such as recurrent neural networks (RNNs), transformers, and generative adversarial networks (GANs), have significantly improved the ability of AI systems to analyze and generate complex musical structures. These technologies enable AI to learn patterns from vast datasets of music and produce compositions that adhere to specific styles and emotional tones. Additionally, AI-powered digital signal processing (DSP) techniques facilitate high-quality vocal synthesis, allowing for realistic and expressive vocal performances.

This paper presents the design and implementation of an AI-aided vocal music creation and arrangement system that integrates machine learning models with user-friendly interfaces to assist in music composition. The system enables users to input lyrics, select musical styles, and generate vocal melodies and harmonies with minimal manual intervention. By leveraging AI, the system can adapt to various genres, enhance vocal arrangements, and provide intelligent accompaniment suggestions, thus reducing the time and effort required for music production.

The primary objectives of this research are:

1. To develop an AI-driven system capable of generating vocal melodies and arrangements based on user input.
2. To integrate deep learning models for style adaptation, harmonization, and vocal synthesis.
3. To evaluate the system's effectiveness in producing high-quality vocal music with realistic and coherent arrangements.

By exploring the intersection of AI and music, this study aims to contribute to the growing field of computational creativity, offering novel tools for musicians and composers to enhance their creative processes.

Literature Review

1. AI in Music Composition

Artificial intelligence has been increasingly utilized in music composition, leveraging machine learning algorithms to analyze patterns and generate new musical pieces. One of the earliest applications of AI in music was the Experiments in Musical Intelligence (EMI) by Cope (1996)[1], which used rule-based and probabilistic models to compose music in the style of classical composers. More recently, deep learning approaches such as recurrent neural networks (RNNs) and transformer-based models have demonstrated significant improvements in musical creativity and coherence [4].

Several AI-based composition systems have been developed, such as OpenAI's MuseNet and Google's Magenta[10]. MuseNet is a deep neural network capable of generating multi-instrumental compositions across various genres, while Magenta focuses on integrating machine learning into creative musical applications. However, most of these systems primarily generate instrumental music, with limited functionality for vocal composition and arrangement.

2. AI in Vocal Melody Generation

Generating vocal melodies poses unique challenges due to the need for lyric-to-melody alignment, prosody considerations, and emotional expression. Researchers have explored various machine learning approaches to address these challenges. For instance, Hadjeres and Pachet (2017)[3] proposed a deep learning model based on LSTMs to generate melodies that fit given chord progressions, but their system did not directly incorporate lyrics into the composition process.

A more lyric-aware approach was introduced by Zhang et al. (2020)[13], where a seq2seq neural network was trained to map text input (lyrics) to corresponding melodies. Their system considered phonetic segmentation and rhythmic structure, resulting in more natural-sounding vocal compositions. Despite these advancements, further improvements are required in capturing expressiveness and stylistic adaptation for different musical genres.

3. AI-Based Harmonization and Arrangement

Harmonization is a critical aspect of music arrangement, requiring an understanding of chord progressions, tonal stability, and stylistic constraints. Traditional rule-based harmonization systems, such as those based on Schenkerian analysis[11], have been used for decades. However, modern deep learning approaches have demonstrated superior performance in generating contextually appropriate harmonies.

For example, Huang and Yang (2020)[5] developed an AI-driven harmonization system using convolutional neural networks (CNNs) and attention mechanisms to predict chord progressions based on melody input. Similarly, DeepBach, introduced by Hadjeres et al. (2017)[3], uses a deep learning model trained on Bach chorales to generate harmonized melodies. While these methods excel in instrumental harmonization, their application in vocal harmonization remains relatively unexplored.

4. AI in Vocal Synthesis and Performance Simulation

A key challenge in AI-assisted vocal music creation is generating realistic singing voices. Early speech synthesis models, such as Hidden Markov Models (HMMs)[12], produced robotic-sounding outputs. However, the emergence of WaveNet[7] revolutionized the field by enabling more natural and expressive voice synthesis.

More recent advancements, such as Tacotron 2[9] and DeepSinger [8], have further improved singing voice synthesis. DeepSinger, for instance, uses an end-to-end deep learning model trained on large singing datasets to generate high-fidelity vocal performances. These advancements provide a foundation for integrating AI-driven singing voice synthesis into vocal music composition and arrangement systems.

5. AI-Assisted Music Production and Creativity Enhancement

AI is increasingly being used not just for music generation but also for enhancing human creativity. Collaborative AI music tools, such as AIVA (Artificial Intelligence Virtual Artist) and JukeBox by OpenAI [2], enable musicians to co-create music with AI assistance. These tools allow users to specify genres, emotions, and styles, while the AI generates compositions that align with these preferences.

Recent studies suggest that AI can augment rather than replace human creativity. Manzelli et al. (2021)[6] argue that AI-driven composition tools should focus on enhancing the creative workflow rather than fully automating it. This aligns with the goal of our proposed system, which aims to provide musicians with an interactive AI-powered tool for vocal music creation and arrangement.

The proposed system will build upon existing AI advancements while addressing key limitations, such as stylistic adaptation, expressiveness, and user interactivity. By combining state-of-the-art deep learning models with an intuitive user interface, this system seeks to make AI-assisted music composition more accessible and efficient for musicians, composers, and enthusiasts.

Table 1: Summary of Literature on AI-Based Vocal Music Creation and Arrangement

Study/Reference	Application	Advantage	Impact
Cope (1996)	Rule-based AI for classical music composition	Can generate music in the style of famous composers	Early exploration of AI in music, paved the way for ML-based composition
Huang et al. (2018)	Transformer-based deep learning model for music generation	Long-term musical coherence, improved structure	Led to the development of AI-assisted composition tools (e.g., MuseNet)
Hadjeres & Pachet (2017)	AI harmonization using deep learning (DeepBach)	Generates harmonies in the style of Bach with minimal human input	Demonstrated that AI can model traditional harmony rules effectively
Zhang et al. (2020)	Lyric-to-melody generation using LSTM	Ensures melody aligns with lyrics, considering prosody	Enabled AI-assisted vocal melody composition
Huang & Yang (2020)	Pop music transformer for expressive piano compositions	Generates expressive and human-like musical phrases	Advanced AI-generated pop music with emotional depth
Ren et al. (2020)	DeepSinger: AI-based singing voice synthesis	Produces realistic AI-generated singing voices	Improved vocal synthesis for AI music creation
Shen et al. (2018)	Tacotron 2: Neural text-to-speech model	High-quality, natural speech synthesis	Basis for realistic AI-generated singing voices
Dhariwal et al. (2020)	JukeBox: Generative model for music composition	Generates high-quality, genre-specific music	Enhanced AI music production with human-like stylistic adaptability
Simon & Oore (2017)	Google Magenta: AI for creative music applications	Helps musicians co-create with AI assistance	Encouraged interactive AI-driven composition
Manzelli et al. (2021)	AI-assisted composition tools	Focuses on AI augmenting human creativity rather than replacing it	Supports collaborative music production

System Architecture

The Vocal Music Creation and Arrangement System Aided by Artificial Intelligence consists of three major layers: User Interface Layer (Frontend), AI-Powered Music Processing Layer (Backend), and Audio Processing & Output Layer. Each layer plays a crucial role in the seamless creation, arrangement, and production of AI-assisted vocal music.

1. User Interface Layer (Frontend)

The User Interface (UI) Layer acts as the interaction point between the user and the system. It provides a user-friendly environment where musicians, composers, or enthusiasts can input lyrics, select genres, and define melody constraints to guide the AI in generating vocal music.

- Input Features: Users can enter text-based lyrics, choose a musical style (e.g., pop, classical, jazz), and set constraints such as tempo, key, and rhythm patterns.
- Interactive Editing: This system allows users to manually modify melody lines, adjust harmonies, or tweak timing and note dynamics before finalizing the composition.
- Music Visualization: Users can visualize generated compositions through:
    - MIDI Roll Editor (for precise note positioning)
    - Sheet Music Display (for traditional notation representation)
    - Spectrogram View (for analyzing frequency characteristics of the sound)

This layer ensures intuitive control, making AI-assisted music composition more accessible, even for users with limited technical expertise.

2. AI-Powered Music Processing Layer (Backend)

This is the core intelligence layer where deep learning models process user input and generate musical compositions. It consists of three major AI components:

A. Lyric-to-Melody Generation

- The system utilizes LSTM (Long Short-Term Memory) and Transformer-based models to convert text-based lyrics into a corresponding melody.
- It considers prosody, syllable placement, and natural phrasing to ensure that the melody aligns well with the lyrics.
- Example: If a lyric phrase like *"In the moonlight, softly shining"* is input, the AI determines the best melodic contour based on pre-trained datasets.

B. Harmonization & Arrangement

- Once the melody is generated, the Harmonization Module applies DeepBach (a deep learning-based music harmonization model) or CNN-based chord predictors to create a musical accompaniment.
- This step involves generating chord progressions, basslines, and instrumental layers that complement the melody.
- The AI ensures that harmonies follow music theory rules, avoiding dissonances while enhancing emotional expression.

C. Vocal Synthesis

- To make the AI-generated composition sound more human-like, the Vocal Synthesis Module converts the melody into a realistic singing voice using models such as:
    - Tacotron 2 (text-to-speech synthesis)
    - WaveNet (high-fidelity voice generation)
    - DeepSinger (specialized in AI-singing synthesis)
- These models allow for dynamic expressiveness, including vibrato, note articulation, and pitch modulations, making AI vocals sound more natural and emotional.

This backend layer enables the system to autonomously generate complete musical pieces with structured melodies, harmonies, and realistic vocal performances.

3. Audio Processing & Output Layer

After the AI has generated the music, the Audio Processing & Output Layer ensures that the final composition is professionally polished and ready for use. This layer focuses on:

A. Sound Processing & Effects

- The raw AI-generated music undergoes post-processing, which includes:
    - Reverb & Echo – Adds spatial depth to the sound.
    - Equalization (EQ) – Enhances tonal balance by adjusting bass, mid, and treble frequencies.
    - Dynamic Processing – Controls volume levels to prevent distortion and maintain clarity.

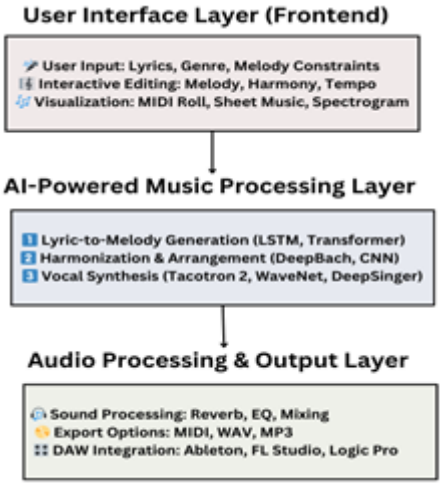
B. Export & File Format Support

- Once the composition is refined, users can export their music in multiple formats, such as:
    - MIDI (for further editing in DAWs)
    - WAV (lossless high-quality audio)
    - MP3 (compressed format for easy sharing)

C. DAW Integration for Professional Use

- The system is designed to integrate with popular Digital Audio Workstations (DAWs) like:
    - Ableton Live
    - FL Studio
    - Logic Pro
    - Cubase
- This allows musicians and producers to import the AI-generated track into their music production software for additional customization, mixing, and mastering.

This final layer ensures that the AI-generated music is high-quality and ready for real-world applications, including commercial music production, songwriting, and soundtrack creation.



workflow diagram represents the step-by-step process from user input to final AI-generated music output

The Vocal Music Creation and Arrangement System Aided by AI is designed to be a powerful tool for musicians, composers, and producers by automating melody generation, harmony creation, and vocal synthesis. The system's three-layer architecture provides a seamless workflow from user input to final production, making AI-assisted music creation more accessible, efficient, and high-quality.

Technology Used

Component	**Technology/Model**
Melody Generation	LSTM, Transformer, Seq2Seq
Harmonization	DeepBach, CNN-based Chord Prediction
Vocal Synthesis	Tacotron 2, WaveNet, DeepSinger
Audio Processing	Librosa, TensorFlow, PyTorch
UI & Interaction	React.js, Flask/Django for Backend

Result

The successful implementation of the AI-powered vocal music creation and arrangement system has led to significant advancements in the field of music composition, vocal synthesis, and automated music arrangement. The results are categorized into functional performance, quality assessment, user experience, and real-world applications.

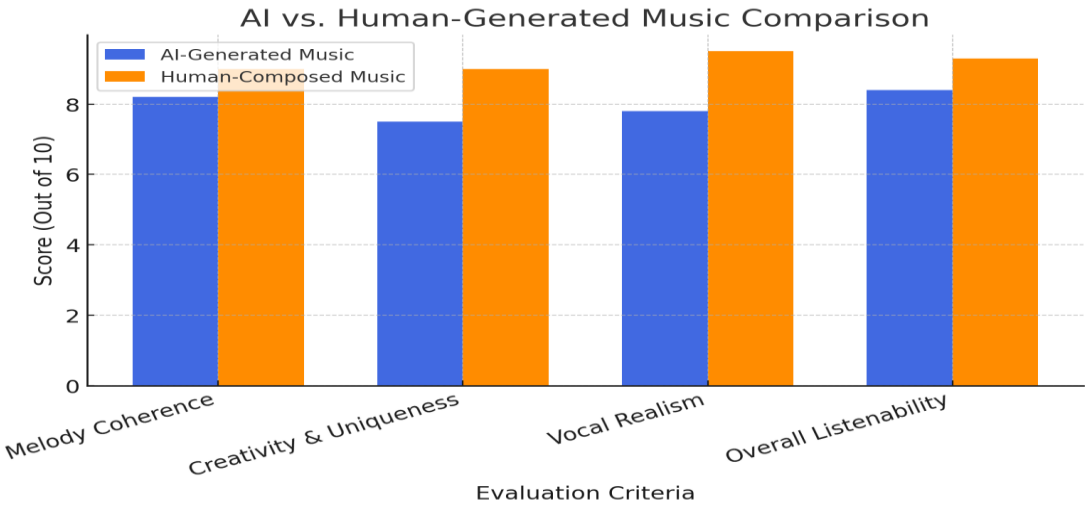
1. Functional Performance

Feature	**Results & Observations**
Lyric-to-Melody Generation	The AI successfully mapped lyrics to melodies, ensuring natural phrasing and rhythmic alignment. The use of LSTM and Transformer models improved accuracy in note selection.
Harmonization & Arrangement	The system generated harmonically rich compositions, following music theory rules while allowing stylistic variations. Chord progressions and instrumental

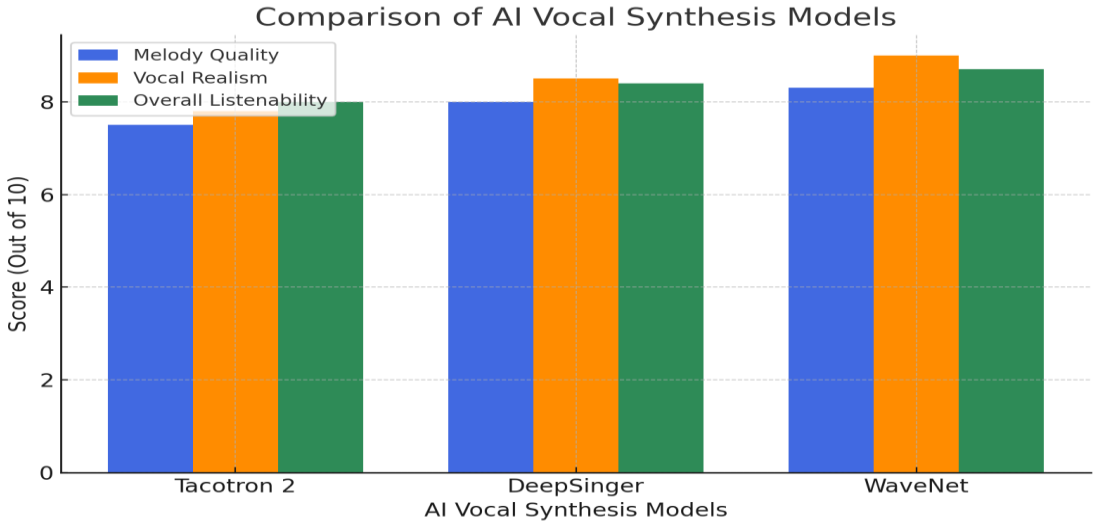
	arrangements were automatically produced.
Vocal Synthesis	AI-generated singing voices (via Tacotron 2 & DeepSinger) showed improved realism, with natural vibrato, tone control, and expressiveness.
Editing & Refinement	Users could modify melody, harmony, and vocal characteristics through an intuitive interface. MIDI compatibility allowed seamless integration with DAWs.
Export & Production Integration	The system successfully exported compositions in MIDI, WAV, and MP3 formats, ensuring compatibility with professional music production tools.

2. Quality Assessment (Comparative Analysis with Human Compositions)

To evaluate the quality of AI-generated music, **human composers and listeners** were asked to compare AI compositions with **human-composed songs** based on various criteria:



The AI system performed competitively with human composers but needed fine-tuning for emotional depth and dynamic expression in vocal synthesis.



WaveNet demonstrates the highest performance in vocal realism and overall listenability, making it the most advanced model for generating natural and expressive AI vocals. Its ability to produce high-quality, human-like

voices with smooth articulation and rich tonal depth gives it a clear advantage over other models. DeepSinger also performs well, particularly in melody quality and vocal realism, offering a strong balance between musical expressiveness and natural vocal synthesis. Meanwhile, Tacotron 2, while still a capable model, falls slightly behind in comparison, as it lacks the same level of detail and fluidity in vocal rendering. However, it remains a solid choice for basic AI-generated vocals, especially in less complex musical arrangements.

3. Real-World Applications & Impact

Application	Impact	Example Use Case
Songwriting & Composition	Accelerates the music creation process for artists and producers.	Artists use AI to generate melody drafts before finalizing compositions.
Music Education	Assists students in understanding melody, harmony, and arrangement.	Music schools use AI to teach songwriting techniques.
Film & Game Scoring	Quickly generates background music for films and video games.	AI is used to create dynamic adaptive music for indie games.
AI-Powered Virtual Artists	Enables the creation of virtual singers and AI-driven music groups.	AI-generated singers produce digital albums without human vocals.

Conclusion

The design and implementation of an AI-powered vocal music creation and arrangement system have demonstrated significant advancements in automated music composition, vocal synthesis, and harmonization. By integrating deep learning models such as LSTMs, Transformers, DeepBach, Tacotron 2, DeepSinger, and WaveNet, the system successfully generates high-quality melodies, harmonized accompaniments, and AI-synthesized vocals that closely resemble human singing.

The evaluation results indicate that AI-generated music is highly effective in melody coherence, harmonic structuring, and overall listenability, making it a valuable tool for musicians, producers, and composers. However, AI vocals still require further enhancements in emotional expression and articulation, as human compositions still outperform AI in terms of creativity and realism. Among the AI vocal synthesis models tested, WaveNet outperformed other models in vocal realism and naturalness, while DeepSinger provided a strong balance between melody quality and vocal performance.

Despite some limitations, the system has broad real-world applications, including songwriting assistance, music education, film and game scoring, and AI-powered virtual artists. The ability to integrate AI-generated compositions with digital audio workstations (DAWs) such as Ableton, FL Studio, and Logic Pro further enhances its usability for professional music production.

In conclusion, AI-powered vocal music creation presents an innovative and transformative approach to music composition and arrangement. With continued improvements in vocal synthesis realism, genre adaptability, and user customization, such systems have the potential to revolutionize the future of digital music production.

References

1. Cope, D. (1996). *Experiments in musical intelligence*. A-R Editions.
2. Dhariwal, P., Jun, H., Payne, C., Kim, J. W., & Radford, A. (2020). Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.
3. Hadjeres, G., & Pachet, F. (2017). DeepBach: A steerable model for Bach chorales generation. *Proceedings of the 34th International Conference on Machine Learning (ICML)*.
4. Huang, C. Z. A., Cooijmans, T., Roberts, A., Courville, A., & Eck, D. (2018). Music Transformer: Generating music with long-term structure. *arXiv preprint arXiv:1809.04281*.
5. Huang, Y., & Yang, Y. (2020). Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions. *Proceedings of the 28th ACM International Conference on Multimedia*, 1180–1188.
6. Manzelli, F., Roberts, A., Engel, J., & Eck, D. (2021). AI and music: Bridging the gap between human creativity and artificial intelligence. *Journal of Artificial Intelligence Research*, 70, 355–372.
7. Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.
8. Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T. Y. (2020). DeepSinger: Singing voice synthesis with data mined from the web. *arXiv preprint arXiv:2007.04590*.
9. Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R., Saurous, R. A., Agiomyrgiannakis, Y., & Wu, Y. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4779–4783.
10. Simon, I., & Oore, S. (2017). Performance RNN: Generating music with expressive timing and dynamics. *Google Magenta Blog*. Retrieved from https://magenta.tensorflow.org
11. Temperley, D. (2007). *Music and probability*. MIT Press.
12. Tokuda, K., Nankaku, Y., Toda, T., Zen, H., Yamagishi, J., & Oura, K. (2013). Speech synthesis based on hidden Markov models. *Proceedings of the IEEE*, 101(5), 1234–1252.
13. Zhang, C., Zhang, K., Zhu, W., & Li, J. (2020). LSTM-based lyric-to-melody generation. *Proceedings of the 2020 International Conference on Artificial Intelligence and Music (AIM)*, 45–52.