Choral Music Generation: A Deep Hybrid Learning Approach

Capitol Technology University

Ph.D. Dissertation Defense

Daniel Szelogowski

Chair: Dr. Susan Goodwin

Committee: Dr. Paul Bodily

Date: March 20, 2024

CAPITOL
Technology University

Statement of the Problem

- Despite significant advancements in AI and machine learning for music generation, there remains a notable gap in the creation of realistic, AI-generated classical music, particularly in multi-voiced compositions like choral music, where achieving a complete, harmonically and structurally coherent score that captures the intricacies and nuances of classical choral compositions demands a high level of musical complexity and understanding.

CAPITOL Technology University

Rationale for the Study

- The study aims to explore and validate the use of machine learning and deep learning architectures for generating musically realistic four-part choral music, thereby advancing AI technology in music composition and deepening the understanding of AI's role in artistic expression.
- By examining existing research, assessing the capability of current AI architectures to produce compositions akin to human creation through a novel benchmark model, and analyzing perceptual differences in AI-generated music among individuals with varied musical expertise, the study seeks to bridge the technological gap in AI music generation and explore its potential for creative enhancement in the arts.

CAPITOL Technology University

Purpose of the Study

- This research offers novels insights into AI's role in creative expression and lays the groundwork for subsequent innovations in AI-driven music creation, examining the perceptual effects of the project's model and its capability to produce compositions that are indiscernible from human-composed music to both musicians and non-musicians.
- The study is thus intended for readers with a strong background in computer science and statistical analysis, and at least elementary knowledge of music theory.

CAPITOL Technology University

Research Questions

1. Is there a current Machine Learning or Deep Learning architecture that is feasible for the generation of musically realistic four-part (SATB) choral music, and does such a training dataset currently exist?

2. Can a hybrid Artificial Intelligence model generate realistic choral music to the degree that the general public would struggle to differentiate from genuine human compositions?

3. Do individuals of varying musical backgrounds perceive AI-generated classical music differently than non-musicians?

CAPITOL
Technology University

Literature Overview

1. Holtzman (1981) explores using **generative grammars** for realistic music composition, overcoming early algorithmic challenges.
2. Van Der Merwe and Schulze's (2011) SuperWillow system uses **Markov Models** and **probabilistic automata** to compose music that mimics human compositions.
3. Buys (2011) presents "BACCHUS," leveraging large datasets and **finite-state transducers** to improve generative music **Markov models'** structure and style imitation.
4. Hadjeres et al. (2017) introduce a graphical model (DeepBach) using a **two-RNN/two-FNN ensemble network with pseudo-Gibbs sampling** to generate Bach-style chorales, allowing user-steered composition through a MuseScore plugin.
5. Mao et al. (2018) present an end-to-end generative **Biaxial LSTM** model (DeepJ) enabling music composition conditioned on specific composer styles, offering tunable parameters for style-specific music creation.

CAPITOL
Technology University

Literature Overview

6. Yanchenko and Mukherjee (2018) explore the use of **state space models** (particularly **Hidden Markov Models**) for composing classical piano pieces with consonant harmonies, noting limitations in melodic progression.
7. Huang (2019)* examines how deep learning enhances musical creativity through generation, recommendation, and control, offering novel tools like **Music Transformer** to make composition more accessible to both novices and musicians.
8. Herremans and Chew (2019) introduce MorpheuS, a **Variable Neighborhood Search** system that generates polyphonic music with structured tension profiles and repeated patterns, addressing the challenge of long-term musical coherence.
9. Carnovalini and Rodà (2020) provide a comprehensive survey on **Computational Creativity** and **Music Generation Systems**, outlining the current state of the art, the challenges ahead, and potential directions for future research in the field.
10. Caren (2020) introduces TRoco, a novel **quotient-based algorithm** that composes music using jazz theory to match desired musical tension/release contours.

CAPITOL Technology University

Literature Overview

11. Micchi et al. (2021) explore human/AI co-creativity in songwriting, detailing the AI-assisted creation of a song using an **LSTM-based model**, highlighting AI's role in enhancing human composition (i.e., AI as "suggestion" vs. as "automation").
12. Naruse et al. (2022) propose a method for pop music generation allowing control over phrase lengths using Transformer-based **Auto-Regressive models**, introducing "PHRASE" and "BAR COUNTDOWN" events for detailed piece structure control and natural endings.
13. Liu et al. (2022) introduce the **Auto-Regressive model** SymphonyNet, utilizing a novel **Music Byte Pair Encoding** and architecture for generating complex, harmonious symphonic music with automatic orchestration learning.
14. Lu et al. (2022) present MeloForm, blending **expert systems with (Transformer-based) neural networks** to generate rich melodies adhering to specific musical forms, achieving high accuracy and quality without labeled form data.
15. Neves et al. (2022) propose a **Transformer-GAN** model to generate music reflecting human sentiment, addressing interaction and coherence challenges in automatic music generation.

CAPITOL Technology University

Research Methodology/Design

- The research method used in this study combined quantitative and qualitative approaches, focusing on applying machine learning to create choral music and evaluating these compositions through surveys.
- This mixed-method allowed for an examination of AI's capability in music generation and an assessment of human reactions to AI-created music.
  - The quantitative part involved training a novel deep learning model (**Choral-GTN**) and analyzing its outputs, while the qualitative part used a survey to assess public perception of AI-generated versus human-composed music.
  - The **Choral-Generative Transformer Network (GTN)** system uses a multi-layer Transformer-based model (the GTN), a MIDI generator callback interface, and a rule-based post-processing system (all written in the Python programming language).
    - The model was trained on a novel 1,000 MIDI dataset, "**CHORAL**," curated by the researcher.
      - **C**horal **H**armony **O**ptimized **R**epository for **A**I **L**earning

CAPITOL Technology University

Research Methodology/Design

- The research design consisted of two phases:
  1. Developing and training the Choral-GTN model on the CHORAL dataset.
  2. Evaluating the AI-generated music with a survey.
- The survey included a musical Turing test to collect diverse listener perceptions.
  - This design facilitated a comprehensive study of AI's ability to produce realistic choral music and how different audiences perceive it.
  - It aligned with the study's goals to demonstrate AI's potential in music creation and explore its perceptual impact.

CAPITOL Technology University

Research Methodology/Design

- Other research methods or designs were not used because they might not have fully captured the study's aims.
  - The chosen approach allowed for a detailed exploration of AI-generated music's technical and perceptual aspects.
  - Other methods might have missed the depth of analysis needed to understand AI's role in music generation and its acceptance by listeners.
- The focus was on balancing AI's creative capabilities with insights into human responses, guiding the selection of this methodology over others.

CAPITOL Technology University

Data Collection

- The study targeted a wide audience, including music teachers/professionals, students, amateurs, and general enthusiasts.
  - This diversity aimed to assess AI music's perception across different levels of musical experience.
  - The sampling approach combined convenience and purposive methods.
    - It included fellow teaching staff, university colleagues, and Prolific.com participants.
    - This method suited the study's exploratory nature, focusing on varied music perceptions.
- Data collection involved AI-generated music samples and an online survey.
  - The GTN model, trained on the CHORAL dataset, produced these samples.
  - The survey, conducted via Pollfish.com and Prolific, collected diverse participant responses.
  - This approach aimed to understand reactions to AI-composed choral music comprehensively.

CAPITOL Technology University

Data Analysis – Instrumentation

- The instrumentation for data analysis encompassed a comprehensive suite of statistical tests tailored to assess various aspects of the AI-generated choral music's perception.
  - A power analysis was conducted to determine the sample size's adequacy, ensuring the reliability of findings.
  - Chi-squared (χ^2) analysis was utilized to identify significant differences in the ability of respondents to distinguish between AI-generated and human-composed music.
  - The Kruskal-Wallis test evaluated variations in response distributions across different levels of musical expertise.
  - Post-hoc analyses, including additional χ^2 tests (larger groups) and Fisher's Exact tests (smaller groups), were performed to further explore differences in perception across demographic groups and their association with opinions on music origin.

CAPITOL Technology University

Data Analysis – Technique

- The analysis technique involved applying several statistical tests to the collected data systematically.
  - Initially, the power analysis informed the reliability of the study's sample size to achieve statistical significance.
  - Subsequently, the χ^2 test assessed the primary hypothesis concerning the recognition of music origin, offering insights into overall trends.
  - The Kruskal-Wallis test then pinpointed differences in musical perception among varying expertise levels, highlighting the model's broad appeal or specific biases.
  - Finally, post-hoc tests delved into demographic-specific responses, clarifying how different listener groups relate to AI-generated music.
- All tests were evaluated for significance where $p < 0.05$.

CAPITOL Technology University

Data Analysis – Technique

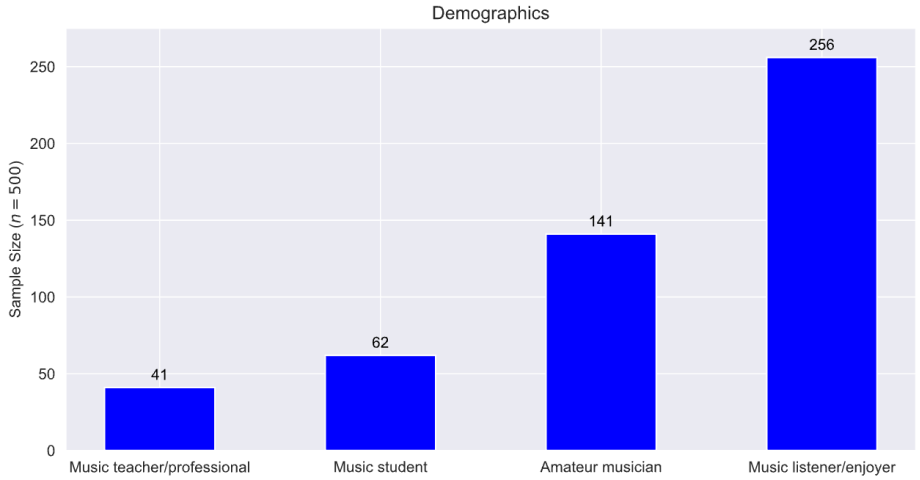

Figure 5.3: Survey demographics by count

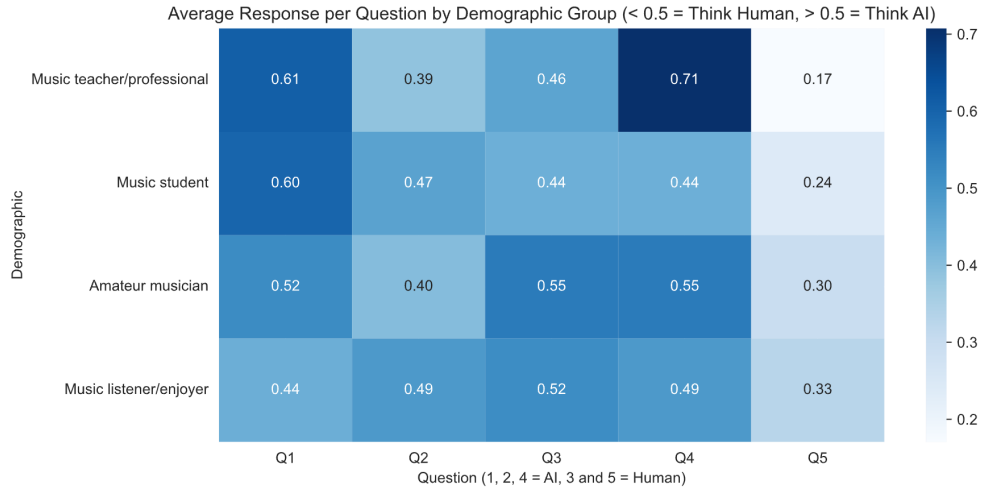

Figure 5.6: Heatmap of mean response distribution by demographic group

CAPITOL Technology University

Data Analysis – Reliability and Validity

- This study aimed for high internal validity by standardizing music sample presentation and survey formats.
  - However, the subjective evaluation of music introduces inherent variability in responses, impacting reliability.
- External validity is bolstered by diverse participant backgrounds but is limited to classical choral music and specific AI models.
  - Efforts to manage errors within and between samples included controlled variables, subjective measures, and randomized sample order, though complete elimination of bias is challenging in subjective assessments.

CAPITOL Technology University

Findings

- Initial χ^2 analysis revealed that the AI model effectively produced compositions indistinguishable from human-created music, with notable perception uniformity across demographic groups, except for one human-composed sample identified significantly as such.
- Kruskal-Wallis tests highlighted variations in distinguishing AI from human compositions, influenced by musical background – especially in music samples from the weaker models, suggesting musical training affects perception.
- Post-hoc analyses further validated these findings, showing no significant perception differences in larger demographic groups across most samples, and indicating nuances detectable by musically trained listeners in certain samples, particularly from the least effective model.
- These results underscore the AI model's capability in mimicking human compositions and the nuanced influence of musical expertise on music perception, reinforcing the model's potential in music composition and the complex interplay of musical training in recognizing AI-generated music.

CAPITOL Technology University

Findings – Answers to Research Questions

- The study explored AI's capability in generating realistic SATB choral music, its recognition by the public, and perception differences among varied musical backgrounds.
  1. Findings confirm the feasibility of AI architectures for creating realistic choral music, evidenced by the hybrid AI model's successful deployment on a curated dataset.
  2. This model convincingly mimicked SATB music's complexity, demonstrating advancements in deep learning and dataset availability.
- The survey indicated the general public, with diverse musical backgrounds, often could not reliably distinguish AI-generated music from human compositions, highlighting the model's proficiency.
  3. However, nuanced perception differences emerged, especially among those with more advanced musical training, suggesting musical expertise influences discernment of AI compositions.
- These findings affirm AI's potential in choral music, highlighting its ability to emulate human creativity and the nuanced role of musical training in perceiving AI-generated music.

CAPITOL Technology University

Conclusions

- This study assessed a hybrid AI model's ability to generate SATB choral music, comparing it to human compositions.
- Analysis involved diverse participants, enhancing the findings' reliability
  - Statistical tests showed uniform perception across demographics, affirming the AI's effectiveness in generating realistic choral compositions.
  - Notably, certain musical nuances were more apparent to trained musicians, hinting at the AI's sophisticated output.
- Overall, the GTN model produced music largely indistinguishable to varied listeners, demonstrating AI's creative potential in music composition.

CAPITOL Technology University

Contribution to the Body of Knowledge

- This dissertation addresses a critical gap in AI-generated choral music, contributing a specialized "CHORAL" dataset and the novel "Choral-GTN" system.
- It expands the body of knowledge by blending advanced machine learning systems with music composition principles, achieving realistic choral outputs challenging for listeners to distinguish from human compositions.
- The study's detailed literature review and comprehensive analysis of musical perception among various backgrounds underscore the intricate relationship between AI creativity and human expertise, enhancing understanding in AI's application to musicology.

CAPITOL Technology University

Recommendations for Future Research

- Future research should explore expanding the Choral-GTN system and CHORAL dataset for broader applications, such as music education and AI-assisted composition tools.
  - Opportunities include extending the model to support more voice parts, integrating instrumental accompaniment, and developing lyric generation capabilities (e.g., for complete choral compositions).
  - The system could be adapted for music composition software (such as a "copilot" system), enhancing educational and creative processes.
  - Research could also delve into incorporating emotional or thematic elements to produce contextually richer music for various storytelling mediums.
- Lastly, this study sets a foundation for examining public perception of AI-generated music, suggesting further investigation into the ethics and recognition of AI compositions.

CAPITOL
Technology University

Thank you!

Questions?