**Table of Contents**

[**1. Dmitry Beresnev, Vsevolod Klyushev 2**](#_s5biyz53y5k2)

[2. Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning 2](#_s2osv10b8su)

[3. Canese, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., & Spanò, S. (2021). Multi-agent reinforcement learning: A review of challenges and applications 2](#_8pxlgmogu0j0)

[4. Li, S.,Wu, Y.,Cui, X.,Dong,H.,Fang, F.,Russell, S.(2019, July).   
Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient 2](#_esragj5za60z)

[5. M. Wen et al., (2022). Multi-Agent Reinforcement Learning is a Sequence Modeling Problem 2](#_aao93mrjuxil)

[**2. David Daniel, Danil Afonchikov, Ayhem Bouabid 2**](#_gjdgxs)

[1. Shih, Yi-Jen, et al. "Theme transformer: Symbolic music generation with theme-conditioned transformer." IEEE Transactions on Multimedia (2022) 2](#_rzpinw4w5eoo)

[2. Hsiao, Wen-Yi, et al. "Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 1. 2021 2](#_spfrnxrg7g6a)

[3. Wang, Ziyu, and Gus Xia. “MuseBERT: Pre-training Music Representation  
for Music Understanding and Controllable Generation” ISMIR. 2021. 2](#_30j0zll)

[4. Sulun, Serkan, Matthew EP Davies, and Paula Viana. "Symbolic music generation conditioned on continuous-valued emotions." IEEE Access 10 (2022): 44617-44626 4](#_1t3h5sf)

[6. Qin, Yang, et al. "Score Images as a Modality: Enhancing Symbolic Music Understanding through Large-Scale Multimodal Pre-Training." Sensors 2024, 24, 5017. 6](#_lnxbz9)

# 1. Dmitry Beresnev, Vsevolod Klyushev

## 2. Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning

## 3. Canese, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., & Spanò, S. (2021). Multi-agent reinforcement learning: A review of challenges and applications

## 4. Li, S.,Wu, Y.,Cui, X.,Dong,H.,Fang, F.,Russell, S.(2019, July). Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient

## 5. M. Wen et al., (2022). Multi-Agent Reinforcement Learning is a Sequence Modeling Problem

# 2. David Daniel, Danil Afonchikov, Ayhem Bouabid

## Shih, Yi-Jen, et al. "Theme transformer: Symbolic music generation with theme-conditioned transformer." IEEE Transactions on Multimedia (2022)

## Hsiao, Wen-Yi, et al. "Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 1. 2021

## 3. Wang, Ziyu, and Gus Xia. “MuseBERT: Pre-training Music Representation for Music Understanding and Controllable Generation” ISMIR. 2021.

### 1. Problem Statement

The literature review effectively identifies the core challenge addressed by the paper: the representation of polyphonic music as a sequence and the embedding of such representations. This is a significant issue in music information retrieval and generation, as traditional models often struggle with the complexity and non-linear nature of polyphonic music. The review succinctly captures the essence of the problem, setting a solid foundation for the subsequent discussion.

### 2. Solution & Key Findings

The review accurately summarizes the proposed solution, MuseBERT, which leverages BERT architecture and introduces a generalized relative position encoding to enhance attention mechanisms for polyphonic music. The assertion that MuseBERT yields results closer to human-composed music than baseline models is a critical finding that underscores the model's effectiveness. However, the review could benefit from a more detailed exploration of the specific methodologies employed in the implementation of MuseBERT, such as the training process, data representation, and the nature of the datasets used. This would provide a clearer understanding of how the model achieves its results.

### 3. Methodology Critique

The review appropriately highlights the mathematical foundation supporting the model's effectiveness, as well as the use of multiple baseline comparisons and ablation studies. The inclusion of both objective and subjective evaluations is a notable strength, as it provides a comprehensive assessment of the model's performance. However, the review could delve deeper into the implications of these methodologies. For instance, a discussion on the robustness of the subjective evaluation process and the criteria used for selecting the baselines would enhance the critique. Additionally, the review could address potential biases in the subjective evaluation, particularly regarding the musical backgrounds of the evaluators.

### 4. Strengths

The strengths identified in the review, such as the dual approach to evaluation and the learned melody embeddings, are well-articulated. The mention of subjective scores being compared to training data as a baseline is particularly relevant, as it contextualizes the model's performance. However, the review could further emphasize the innovative aspects of MuseBERT's architecture and its potential implications for future research in music generation and analysis.

### 5. Weaknesses

The review correctly points out the weaknesses, including the reliance on generic metrics for subjective evaluation and the use of low-level music objective metrics. While these critiques are valid, the review could expand on the potential impact of these weaknesses on the overall findings. For instance, discussing how the choice of metrics may limit the model's applicability in real-world scenarios or its ability to capture the nuances of musical expression would provide a more critical perspective. Furthermore, the review could explore alternative evaluation metrics that could enhance the assessment of musical quality.

**Literature Review Evaluation Score: 4.5/5**

**Justification:**

The review effectively identifies the core problem of representing polyphonic music and succinctly summarizes the innovative solution proposed by the authors.

The strengths highlighted, such as the dual approach to evaluation (both objective and subjective) and the learned melody embeddings, reflect a thorough analysis of the paper's contributions. Additionally, the critique of the methodologies used, including the mathematical foundation and the comprehensive evaluation process, indicates a critical engagement with the material.

However, while the review is well-structured and insightful, there are areas for improvement. A deeper exploration of the specific methodologies employed in MuseBERT, as well as a more nuanced discussion of the implications of the identified weaknesses, would enhance the overall depth of the review. Furthermore, suggestions for alternative evaluation metrics could provide a more rounded critique.

Despite these minor shortcomings, the review effectively captures the essence of the paper and provides a solid foundation for understanding MuseBERT's significance in the field of music representation and generation. The clarity of expression, critical engagement, and overall coherence of the review justify a high score of 4.5 out of 5.

## 4. Sulun, Serkan, Matthew EP Davies, and Paula Viana. "Symbolic music generation conditioned on continuous-valued emotions." IEEE Access 10 (2022): 44617-44626

### 1. Problem Statement

The literature review provides a concise introduction to the paper by Sulun et al. (2022), effectively outlining the central problem of traditional conditional music generation models that rely on predefined musical features. The review correctly identifies the emotional aspect of music as a critical element that these models often overlook, setting the stage for the authors' innovative approach that leverages the arousal-valence model to enhance music generation.

### 2. Solution & Key Findings

The review accurately summarizes the authors' solution, which involves the creation of a novel dataset that pairs MIDI music with continuous arousal values sourced from the Spotify Developers API. This dataset serves as a foundation for training a Music Transformer model, which is initially pre-trained for unconditional music generation and subsequently fine-tuned on the newly collected dataset. The comparison of three methodologies for incorporating conditioning signals—discrete tokens, continuous-added tokens, and continuous concatenated tokens—is well-articulated, with the continuous concatenated approach highlighted as the most effective. This section is well-presented, as it succinctly captures the essence of the authors' contributions and findings.

### 3. Methodology Critique

The critique of the methodologies used in the paper is appropriately noted in the review. The reliance on objective evaluation metrics such as negative log-likelihood (NLL) and accuracy is highlighted as a limitation, particularly given the absence of domain-specific metrics that could provide a more nuanced understanding of the model's performance. Furthermore, the lack of a baseline model that employs alternative conditional generation techniques is a significant oversight that the review rightly points out. However, the review could have benefited from a deeper exploration of how these methodological choices impact the overall validity and applicability of the findings.

### 4. Validation and Results

While the review mentions the evaluation metrics used, it lacks a detailed discussion of the results obtained from these evaluations. A more thorough examination of the performance outcomes, including specific numerical results or comparisons between the different methodologies, would enhance the reader's understanding of the effectiveness of the proposed approaches. Additionally, the review could have addressed how the results contribute to the broader field of symbolic music generation and emotional modeling.

### 5. Strengths and Weaknesses

The strengths of the paper, particularly the creation of a new dataset and the pre-training of a generic music generator, are well-articulated in the review. These contributions are indeed significant for advancing research in the field. However, the weaknesses identified—such as the lack of subjective evaluation, absence of a baseline model, and the reliance on non-domain-specific metrics—are crucial points that warrant further discussion. The review could have elaborated on the implications of these weaknesses, particularly regarding the potential impact on the perceived quality and emotional resonance of the generated music.

**Evaluation of the Literature Review: Score 4.5/5**

**Justification:**

**1. Clarity and Structure (5/5):** The literature review is exceptionally well-structured, presenting the information in a clear and coherent manner. It effectively introduces the central problem and outlines the significance of the research, making it easy for readers to follow the authors' arguments and contributions.

**2. Depth of Analysis (4/5):** The review provides a solid analysis of the methodologies employed in the paper. It identifies key strengths, such as the innovative approach to emotional modeling and the creation of a new dataset. While it could delve deeper into some aspects, the analysis is still comprehensive and insightful.

**3. Coverage of Key Aspects (4/5):** The review successfully covers the essential elements of the paper, including the problem statement, proposed solutions, and key findings. It highlights the importance of the emotional aspect in music generation, which is a critical contribution to the field. A bit more detail on the results would enhance this aspect, but overall, it is well-rounded.

**4. Relevance and Significance (5/5):** The review effectively emphasizes the relevance of the paper's contributions to the field of symbolic music generation and emotional modeling. It articulates how the authors' work addresses existing gaps in the literature and advances the understanding of music and emotion, showcasing the significance of their findings.

**5. Critical Perspective (4/5):** The review identifies several weaknesses, such as the lack of subjective evaluation and baseline models, which demonstrates a critical perspective. While it could further explore the implications of these limitations, the acknowledgment of these issues reflects a thoughtful engagement with the paper's content.

Overall, the literature review is a strong and insightful piece that effectively summarizes and critiques the paper by Sulun et al. It highlights the key contributions and findings while maintaining clarity and coherence throughout. The minor areas for improvement do not detract significantly from its overall quality, justifying a high score of 4.5 out of 5.

## 6. Qin, Yang, et al. "Score Images as a Modality: Enhancing Symbolic Music Understanding through Large-Scale Multimodal Pre-Training." Sensors 2024, 24, 5017.

### 1. Problem Statement

The literature review provides a concise introduction to the paper by Qin et al. (2024), effectively outlining the central problem of symbolic music understanding and the limitations of single-modal models. The review highlights the necessity for a more comprehensive approach that integrates both visual and symbolic representations of music, which is a pertinent observation given the complexity of musical structures.

### 2. Solution & Key Findings

The review accurately summarizes the authors' proposed solution, the Score Images as a Modality (SIM) model, and its innovative methodologies, including masked bar-attribute modeling and score-MIDI matching. However, it could benefit from a more detailed exploration of how these methodologies specifically contribute to the model's performance and the theoretical underpinnings that justify their selection. A deeper analysis of the implications of these findings on the broader field of music understanding would enhance the review's depth.

### 3. Methodology Critique

The critique of the methodology is well-articulated, particularly regarding the use of a pre-trained Vision Transformer (ViT) and the single-stream approach. However, the review could further elaborate on the potential limitations of this approach, such as the challenges associated with integrating visual and symbolic data in real-time applications. Additionally, a discussion on the robustness of the model's performance across different datasets would provide a more comprehensive evaluation of its methodological soundness.

### 4. Validation and Results

While the review mentions the lack of evaluation across diverse musical genres and notational styles, it does not sufficiently address the validation methods employed by the authors. A critical examination of the validation process, including the metrics used to assess the model's performance, would provide a clearer picture of the model's effectiveness. Furthermore, discussing the results in relation to existing benchmarks in the field would contextualize the significance of the findings.

### 5. Strengths

The review effectively identifies the strengths of the paper, particularly the integration of score images and the novel pre-training tasks. However, it could further emphasize the potential impact of the curated dual-modality dataset on the model's training efficacy and generalizability. Highlighting how this dataset compares to existing datasets in terms of diversity and comprehensiveness would strengthen the argument regarding the model's robustness.

### 6. Weaknesses

The review rightly points out several weaknesses, including the limited evaluation of the model's performance across various musical genres and the lack of thorough examination of computational efficiency. However, it could also mention the potential implications of these weaknesses on the model's applicability in real-world scenarios, such as music generation and performance. Additionally, suggesting specific areas for future research, such as exploring alternative pre-training tasks or enhancing computational efficiency, would provide constructive feedback for the authors.

**Evaluation of the Literature Review: Score 4.5/5**

**Justification**

1. **Clarity and Structure (5/5)**: The review is exceptionally well-structured, providing a clear introduction to the problem and the proposed solution, making it easy for readers to follow.
2. **Identification of Key Contributions (5/5)**: It effectively highlights the innovative aspects of the SIM model, such as the integration of score images and novel pre-training tasks, crucial for understanding its potential impact.
3. **Critical Perspective (4/5)**: The review offers a balanced critique, acknowledging both strengths and weaknesses of the methodologies used. However, a deeper exploration of certain aspects could enhance this section.
4. **Relevance to the Field (4.5/5)**: By situating the work within the broader context of music information retrieval and AI, the review underscores the significance of the findings, though a more detailed discussion of implications could strengthen this relevance.
5. **Areas for Improvement (4/5)**: While it identifies weaknesses, such as limited evaluation across different musical genres, suggesting specific areas for future research would enhance its overall impact.

In summary, the literature review is a well-crafted and insightful analysis that effectively communicates the importance of the research, justifying a score of 4.5 out of 5 and reflecting its value in the academic discussion surrounding symbolic music understanding.