Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

Theory Structured Harmonic Embeddings for Chord Conditioned Melody Generation

Junnuo Wang

New York University, 70 Washington Sq South, New York, NY, USA

Abstract: *Chord-conditioned melody generation remains limited by coarse harmonic encodings that collapse extended, altered, and modal chords into a few root–quality labels, preventing models from learning nuanced tension and resolution behavior. This paper introduces a theory-structured framework that enriches harmonic conditioning and guides decoding in a modular Transformer architecture. First, a Theory-Structured Harmonic Embedding decomposes each chord into additive Root, Quality, Extension, and Tension components, yielding interpretable sub-embeddings without incurring a combinatorial chord vocabulary. Second, a Harmony-Aware Soft Constrained Decoding scheme adjusts pitch logits at inference time using music-theoretic priors on chord-tone preference, tension validity, non-chord-tone resolution, and scale adherence, controlled by a single constraint-strength parameter. Experiments on the Enhanced Wikifonia Leadsheet Dataset compare a CMT-style baseline, an EC2-VAE model, and three ablation variants. The full model significantly improves Chord Tone Ratio, Tension Correctness, and Non-Chord-Tone Resolution, while maintaining corpus-level pitch and rhythm statistics as measured by MGEval KLD and overlap area. These results demonstrate that explicit harmonic structure and theory-aware decoding jointly yield melodies that are both stylistically faithful and more music-theoretically aligned.*

Keywords: *Symbolic Music, Melody Generation, Chord Conditioning*

1. Introduction

Recent advances in symbolic music generation are largely driven by attention-based Transformer models [1]–[3], which capture long-range dependencies and produce fluent sequences. Chord-conditioned melody generation remains a core task: the model must produce a monophonic line that is rhythmically coherent and harmonically aligned with a chord progression [4]–[6]. Despite architectural progress, current systems still struggle to encode fundamental music-theoretic structure [7], [8], often yielding locally smooth melodies that lack hierarchical rhythmic organization, functional clarity, and convincing tension–resolution behavior.

A key limitation is the oversimplified harmonic input. Most systems condition on a chord root and coarse quality label (e.g., “C major,” “E minor”) [6], [9], [10], collapsing harmonically distinct entities such as extended, altered, or modal chords into a single token. This abstraction discards information about characteristic sevenths, modal colorations, and altered tensions (e.g.,♭9, ♯11) that shape melodic choices. When models must infer these distinctions implicitly, they frequently mishandle chord tones, tensions, and non-chord-tone resolution. In practice, expressive capacity is constrained less by architecture than by the sparsity of the harmonic signal.

This paper proposes a unified framework that enriches harmonic conditioning and guides decoding. A Theory-Structured Harmonic Embedding decomposes each chord into additive Root, Quality, Extension, and Tension components, yielding interpretable sub-embeddings that encode functional roles while avoiding the combinatorial explosion of full chord vocabularies. A Harmony-Aware Soft Constrained Decoding scheme then adjusts pitch-decoder logits at inference time to encourage chord-tone preference, tension validity, and appropriate resolution patterns, controlled by a single scalar that enables post-training interpretability and controllability.

We investigate: (1) whether structured harmonic embeddings improve chord-tone correctness, tension handling, and resolution accuracy; (2) whether soft constraints promote theory-aligned behavior without sacrificing generative diversity. Our contributions are: (a) a structured, interpretable chord-embedding design grounded in music theory; (b) a decoding mechanism that injects harmonic knowledge directly into the generative process; and (c) a comprehensive evaluation that combines theory-aligned

ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

metrics with statistical measures from the MGEval framework.

2. Related Work

2.1 Sequence Models and Controllable Generation

Symbolic music generation is dominated by attention-based sequence models, with the Transformer architecture [1], [2] enabling effective modeling of long-range musical structure through event-based representations [4], [11], [12]. Although recent work explores diffusion models for symbolic music and chord-conditioned melody generation [13], [14] and state-space sequence models for long-range structure [15], the autoregressive Transformer remains the standard for chord-conditioned generation and controlled ablation studies. As model sizes range from millions to billions of parameters [16], [17], efficient conditioning becomes essential. This work emphasizes improving harmonic representation rather than increasing architectural scale.

Controllability is typically achieved through conditional architectures [4], [5], [10], [11]. The Chord-conditioned Melody Transformer (CMT) [4] employs a Chord Encoder and separate Rhythm and Pitch Decoders, demonstrating strong performance with augmented datasets. However, its simplified chord encoding limits expressivity with complex harmonic vocabulary. Latent-variable approaches such as EC2-VAE [18] and hierarchical VAEs for music [19] provide abstract control but require nontrivial disentanglement techniques. By contrast, our method achieves explicit, interpretable control by enriching the harmonic embedding and applying theory-informed logit biasing rather than manipulating latent spaces.

2.2 Music-Theoretic Constraints and Objective Evaluation

Integrating music-theoretic principles has become an important strategy for improving coherence and interpretability in generative models [7]. Hard rules guarantee correctness but often reduce creativity, motivating soft-constraint approaches that bias but do not override the model's learned distribution [20], [21]. Our framework adopts this principle by applying harmonic priors during inference while preserving stylistic fluency.

Evaluating generative music requires metrics beyond token accuracy or loss, which fail to capture multimodal musical validity. We therefore rely on theory-aligned measures [22], Chord Tone Ratio (CTR), Tension Correctness (TC), and Non-Chord Tone Resolution Score (NCTRS), to assess harmonic behavior [6]. To ensure constraints do not distort the stylistic manifold, we use the MGEval framework [23] to compare generated and training MIDI via pitch- and rhythm-based descriptors.

3. Methodology

3.1 Theory-Structured Harmonic Embedding

The central component of our approach is the explicit and fine-grained representation of harmonic function [7]. At each timestep t , the conditioning chord c_t is mapped to a structured, additive embedding e_t^{chord}. This embedding decomposes the chord into interpretable musical factors:

$$e_t^{chord} = e_t^{root} + e_t^{qual} + \sum_{x \in X_t} e_x^{ext} + \sum_{z \in Z_t} e_z^{ten}. \tag{1}$$

This additive formulation is essential. Rather than assigning a unique vector to every possible chord type, an approach that would cause severe data sparsity and a combinatorial explosion [8], the model operates over a vocabulary of functional musical components. The vocabulary consists of 12 root classes, nine chord qualities (e.g., Maj7, Min7, Dom7), three extensions (9, 11, 13), and six common alterations (♭9, ♯9, ♯11, ♭13, etc.) [10], [24], [25].

The symbolic sequence is quantized at 16th-note resolution, and the chord embedding e_t^{chord} is repeated at each timestep to maintain continuous harmonic context. A metrical positional embedding encodes the 16 subdivisions of the bar [26], which is crucial for later soft constraints that operate on strong-beat structure.

3.2 Dataset

The Enhanced Wikifonia Leadsheet Dataset (EWLD), a music lead sheet dataset with more than 5.000 scores [27], is selected for its density of extended and altered chords, making it well-suited for evaluating harmonically aware generative systems. Preprocessing begins with quantization of melody and chord sequences to 16th-note resolution, ensuring rhythmic alignment.

Scores lacking either chord symbols or melodic lines were excluded. To promote functional rather than absolute pitch learning, all songs are transposed into either C major or A minor. The dataset is then segmented into eight-bar windows and divided into 8:1:1 training/validation/test splits.

3.3 Modular Architecture

The system follows a three-module architecture described in prior work [4]: a Chord Encoder, a Rhythm Decoder, and a Pitch Decoder. The Chord Encoder E_{chord} is a Transformer encoder that processes the structured chord embeddings to produce contextual harmonic representations H. The Rhythm Decoder D_{rhythm} is an autoregressive Transformer that predicts onset, sustain, and rest tokens conditioned on H. The Pitch Decoder D_{pitch} is another autoregressive Transformer that predicts pitch tokens while attending to both H and the predicted rhythm sequence.

Training proceeds in two phases. First, E_{chord} and D_{rhythm} are trained jointly. Because rhythm patterns are invariant to key, transposition augmentation across all 12 keys is applied, substantially increasing rhythm-level training data and improving robustness. In the second phase, D_{pitch} is trained with D_{rhythm} frozen. Training remains in the canonical key, enabling melodies to be generated functionally and later transposed to any desired key at inference time.

3.4 Harmony-Aware Soft Constrained Decoding

During inference, the pitch logits produced by D_{pitch} are refined using a set of music-theoretic soft constraints. For each timestep t and candidate pitch p, the model modifies the raw logits $\ell_{t,p}$ as:

$$\tilde{\ell}_{t,p} = \ell_{t,p} + \lambda \Delta_{t,p}, \tag{2}$$

Where

$$\Delta_{t,p} = \sum_{k=1}^{K} \alpha_k f_k(t,p). \tag{3}$$

Here λ is a global constraint-strength parameter, α_k are fixed non-negative weights, and each constraint term $f_k(t,p)$ returns a score in *[0,1]*. C_1 rewards chord tones on metrically strong positions identified by the bar-level positional embedding, and is near zero elsewhere. C_2 penalizes pitches that conflict with the chord's allowed extensions or alterations. C_3 rewards non-chord tones whose subsequent note resolves by step to a chord tone within a short temporal window. C_4 applies a mild penalty to pitches outside the local scale implied by the chord, reducing unnecessary chromaticism while allowing occasional expressive deviations [28]–[30].

4. Experiments

4.1 Experimental Design

All models were trained for 150 epochs on the EWLD training split and evaluated on the held-out test set by generating one eight-bar melody for each test chord progression, yielding 2,950 generated sequences following the CMT protocol [4]. We consider two baselines and three ablation variants. The CMT-style baseline encodes each chord using only its root and generates melodies via unconstrained sampling ($\lambda = 0$) [4]. The EC2-VAE baseline provides a latent-variable comparison model, using variational sampling conditioned on harmonic features [18].

ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

Table 1 Summary of Experimental Model Variants.

Model	Input Representation	Decoding Method
CMT-style Baseline	Root	Standard Sampling
EC2-VAE	Latent Variable	Standard Sampling
A1: Structured Only	Full Structured Input	Standard Sampling
A2: Coarse + Constraints	Root	Soft Constraints
A3: Full Model	Full Structured Input	Soft Constraints

The ablation variants examine the role of structured harmonic input and soft-constraint decoding. A1 uses the full structured harmonic embedding with unconstrained decoding ($\lambda = 0$). A2 applies soft-constraint decoding while retaining the coarse harmonic input of the baseline. A3 represents the full model configuration with both structured input and soft constraints.

4.2 Theory-Driven Objective Metrics

To evaluate the musicological validity of generated melodies, we employ three objective metrics aligned with tonal theory. Chord Tone Ratio (CTR) measures the percentage of note onsets that fall on chord tones [6]. Tension Correctness (TC) measures the proportion of generated tensions or extensions that belong to the set explicitly allowed by the chord symbol. Non-Chord-Tone Resolution Score (NCTRS) quantifies voice-leading quality by measuring how often non-chord tones resolve by step to a chord tone within a short temporal window. Together, these metrics target stability, tension validity, and classical tension–resolution behavior.

4.3 Diversity and Statistical Fidelity

To ensure that the theory-aware soft constraints do not distort corpus-level style or reduce diversity, we evaluate statistical fidelity using the original MGEval framework [23]. For each generated melody and each EWLD test melody, we extract all nine features defined in MGEval: pitch count (PC), pitch class histogram (PCH), pitch class transition matrix (PCTM), pitch range (PR), pitch interval (PI), note count (NC), average inter-onset interval (IOI), note length histogram (NLH), and note length transition matrix (NLTM).

For each feature f , we compute Euclidean distances between all pairs of test melodies $\delta^f(M_{test}, M_{test})$and between test melodies and model outputs $\delta^f(M_{test}, M_{model})$. Following Yang & Lerch [23] and Choi et al. [4], these distance sets are converted to probability density functions, and their similarity is summarized by Kullback–Leibler divergence (KLD) and overlapping area (OA) [23]. Low KLD and high OA indicate that the model reproduces the pitch and rhythm statistics of EWLD for the feature f .

4.4 Constraint-Strength Sensitivity

We analyze the effect of the constraint-strength parameter λ on theory alignment and diversity. Using the trained full model A3, we vary λ only at inference time with $\lambda \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$, and generate one eight-bar melody per test progression.

For each λ, we compute CTR, TC, and NCTRS as in Section 4.2. To assess diversity and style, we report average KLD and OA across the nine MGEval features over the generated corpus. This sensitivity analysis characterizes the trade-off between stronger theory constraints and preserved variety.

5. Results

5.1 Theory-Aligned Harmonic Behavior

Table 2 reports the theory-aligned metrics on the test progressions. The dataset row provides a human reference. The CMT-style baseline slightly exceeds the dataset in CTR but lags behind in TC and NCTRS, while EC2-VAE performs worst on all three measures. Adding Theory-Structured Harmonic Embeddings (A1) yields clear gains over CMT, particularly in TC and NCTRS. Applying only Harmony-Aware Soft Constrained Decoding to coarse input (A2) produces intermediate performance, improving TC and NCTRS relative to CMT with similar CTR. The full model (A3) achieves the best results overall, with the highest CTR, TC, and NCTRS among models and a noticeably smaller gap to the dataset in TC

ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

and NCTRS.

Table 2 Theory-aligned metrics on test progressions.

Model	CTR (%)	TC (%)	NCTRS (%)
Dataset	70.51	80.20	79.60
EC2-VAE	63.15	61.05	57.41
CMT-style Baseline	72.35	69.25	63.55
A1: Structured Only	73.40	75.12	69.35
A2: Coarse + Constraints	72.20	74.80	67.10
A3: Full Model	**75.14**	**77.63**	**72.45**

5.2 Stylistic Fidelity under MGEval

Table 3 shows scalar MGEval features for the test set and all models. All variants stay close to the corpus in PR and IOI. A3 tracks the dataset as closely as the CMT baseline, indicating that the structured embedding and constraints do not distort basic pitch-rhythm statistics.

Table 3 MGEval scalar features (mean ± standard deviation).

Model	PC	PR	PI	NC	IOI
Dataset	8.10 ± 2.41	13.00 ± 3.52	2.50 ± 0.86	30.20 ± 9.80	0.27 ± 0.09
CMT-style Baseline	7.20 ± 1.95	12.10 ± 3.11	2.30 ± 0.60	29.10 ± 9.93	0.27 ± 0.09
EC2-VAE	6.10 ± 2.22	11.80 ± 4.80	3.10 ± 1.41	31.50 ± 13.72	0.29 ± 0.13
A1: Structured Only	7.30 ± 1.83	12.30 ± 3.04	2.40 ± 0.63	29.40 ± 9.60	0.27 ± 0.10
A2: Coarse + Constraints	7.10 ± 1.93	12.00 ± 3.03	2.35 ± 0.60	29.00 ± 9.81	0.27 ± 0.09
A3: Full Model	**7.40 ± 1.85**	**12.60 ± 3.10**	**2.40 ± 0.58**	**29.60 ± 9.72**	**0.27 ± 0.09**

We also compare intra-set and inter-set distance distributions for all nine MGEval features using Kullback–Leibler divergence (KLD) and overlapping area (OA). Table 4 reports averages over the nine features. CMT already models corpus statistics well. EC2-VAE shows much higher average KLD and lower OA. A1 and A2 stay in the same range as CMT, and the full model A3 achieves slightly lower divergence and higher overlap, confirming that theory-aware control does not move the model away from the EWLD manifold.

Table 4 Average KLD and OA.

Model	Avg. KLD (↓)	Avg. OA (↑)
CMT-style Baseline	0.02	0.94
EC2-VAE	0.11	0.85
A1: Structured Only	0.02	0.95
A2: Coarse + Constraints	0.03	0.93
A3: Full Model	**0.01**	**0.96**

5.3 Effect of Constraint Strength

Table 5 summarizes the effect of varying λ for A3. Increasing λ from 0.0 to 0.5 steadily improves theory-aligned metrics, with only small additional gains beyond $\lambda = 0.5$. Average MGEval KLD simultaneously decreases and OA increases, indicating closer alignment with EWLD. For $\lambda \geq 0.75$,, KLD rises slightly and OA falls modestly, suggesting a mild loss of stylistic flexibility. Overall, $\lambda = 0.5$ provides a good balance between theory alignment and corpus-level fidelity and is used in the main experiments.

Table 5 Effect of constraint strength λ for full model A3.

λ	CTR (%)	TC (%)	NCTRS (%)	Avg. KLD (↓)	Avg. OA (↑)
0.00	73.50	75.00	69.00	0.020	0.950
0.25	74.30	76.40	71.00	0.015	0.955
0.50	75.14	77.63	72.45	0.010	0.960
0.75	75.60	78.00	73.00	0.012	0.955
1.00	75.80	78.20	73.20	0.018	0.948

ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

6. Conclusion

This paper tackles the gap between powerful sequence models and weak music-theoretic structure in chord-conditioned melody generation. We showed that coarse root–quality chord encodings limit harmonic control and tension–resolution behavior, and proposed a two-part framework to address this: a Theory-Structured Harmonic Embedding that factorizes chords into Root, Quality, Extension, and Tension components, and a Harmony-Aware Soft Constrained Decoding scheme that reshapes pitch logits using music-theoretic priors under a single tunable constraint parameter.

On the Enhanced Wikifonia Leadsheet Dataset, both components independently improve Chord Tone Ratio, Tension Correctness, and Non-Chord-Tone Resolution over a CMT-style baseline, while the full model achieves the best overall theory-aligned performance and closely tracks corpus pitch–rhythm statistics under MGEval. These results indicate that explicit harmonic structure and soft, theory-informed decoding can produce melodies that are more music-theoretically consistent without compromising stylistic fidelity. Future work includes extending the approach to polyphony, broader genres, and human listening studies to better connect objective metrics with perceived musical quality.

References

[1] Y.-J. Shih, S.-L. Wu, F. Zalkow, M. Müller, and Y.-H. Yang, "Theme Transformer: Symbolic music generation with theme-conditioned Transformer," IEEE Trans. Multimedia, vol. 25, pp. 3495–3508, 2023, doi: 10.1109/TMM.2022.3161851.

[2] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[3] S. Ji, X. Yang, and J. Luo, "A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges," ACM Comput. Surv., vol. 56, no. 1, pp. 1–39, Jan. 2024, doi: 10.1145/3597493.

[4] K. Choi, J. Park, W. Heo, S. Jeon, and J. Park, "Chord conditioned melody generation with Transformer based decoders," IEEE Access, vol. 9, pp. 42071–42080, 2021, doi: 10.1109/ACCESS.2021.3065831.

[5] B. Genchel, A. Pati, and A. Lerch, "Explicitly conditioned melody generation: A case study with interdependent RNNs," arXiv preprint arXiv:1907.05208, Jul. 2019, doi: 10.48550/arXiv.1907.05208.

[6] Y.-C. Yeh et al., "Automatic melody harmonization with triad chords: A comparative study," J. New Music Res., vol. 50, no. 1, pp. 1–15, 2021, doi: 10.1080/09298215.2021.1873392.

[7] S. E. Ni-Hahn, "Machine learning and music theory: Models for hierarchical music generation and analysis," Ph.D. dissertation, Dept. Electr. Comput. Eng., Duke Univ., Durham, NC, USA, 2025.

[8] S.-L. Wu and Y.-H. Yang, "The Jazz Transformer on the front line: Exploring the shortcomings of AI-composed music through quantitative measures," arXiv preprint arXiv:2008.01307, Aug. 2020, doi: 10.48550/arXiv.2008.01307.

[9] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "MidiNet: A convolutional generative adversarial network for symbolic-domain music generation," arXiv preprint arXiv:1703.10847, Jul. 2017, doi: 10.48550/arXiv.1703.10847.

[10] M. Kaliakatsos-Papakostas, D. Makris, K. Soiledis, K.-T. Tsamis, V. Katsouros, and E. Cambouropoulos, "HarmonyTok: Comparing methods for harmony tokenization for machine learning," Information, vol. 16, no. 9, p. 759, Sep. 2025, doi: 10.3390/info16090759.

[11] Y.-S. Huang and Y.-H. Yang, "Pop music Transformer: Beat-based modeling and generation of expressive pop piano compositions," in Proc. 28th ACM Int. Conf. Multimedia, Seattle, WA, USA, Oct. 2020, pp. 1180–1188, doi: 10.1145/3394171.3413671.

[12] H.-W. Dong, K. Chen, S. Dubnov, J. McAuley, and T. Berg-Kirkpatrick, "Multitrack music Transformer," in ICASSP 2023 – IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10094628.

[13] L. Min, J. Jiang, G. Xia, and J. Zhao, "Polyffusion: A diffusion model for polyphonic score generation with internal and external controls," arXiv preprint arXiv:2307.10304, Jul. 2023, doi: 10.48550/arXiv.2307.10304.

[14] S. Li and Y. Sung, "MelodyDiffusion: Chord-conditioned melody generation using a Transformer-based diffusion model," Mathematics, vol. 11, no. 8, p. 1915, 2023, doi: 10.3390/math11081915.

[15] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," arXiv preprint arXiv:2111.00396, 2022, doi: 10.48550/arXiv.2111.00396.

[16] A. Agostinelli et al., "MusicLM: Generating music from text," arXiv preprint arXiv:2301.11325, Jan. 2023, doi: 10.48550/arXiv.2301.11325.

ISSN 2616-5775 Vol. 8, Issue 10: 30-36, DOI: 10.25236/AJCIS.2025.081005

[17] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, "Jukebox: A generative model for music," arXiv preprint arXiv:2005.00341, Apr. 2020, doi: 10.48550/arXiv.2005.00341.

[18] R. Yang, D. Wang, Z. Wang, T. Chen, J. Jiang, and G. Xia, "Deep music analogy via latent representation disentanglement," arXiv preprint arXiv:1906.03626, Oct. 2019, doi: 10.48550/arXiv.1906.03626.

[19] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in Proc. 35th Int. Conf. Machine Learning, Jul. 2018, pp. 4364–4373.

[20] G. Hadjeres, F. Pachet, and F. Nielsen, "DeepBach: A steerable model for Bach chorales generation," in Proc. 34th Int. Conf. Machine Learning, Jul. 2017, pp. 1362–1371.

[21] S. Lattner, M. Grachten, and G. Widmer, "Imposing higher-level structure in polyphonic music generation using convolutional restricted Boltzmann machines and constraints," J. Creat. Music Syst., vol. 2, pp. 1–31, Nov. 2020, doi: 10.3316/informit.668426282761995.

[22] A. C. Salem, M. Shokri, and J. Devaney, "Chord-conditioned melody and bass generation," arXiv preprint arXiv:2511.08755, Nov. 2025, doi: 10.48550/arXiv.2511.08755.

[23] L.-C. Yang and A. Lerch, "On the evaluation of generative models in music," Neural Comput. Appl., vol. 32, no. 9, pp. 4385–4416, 2020, doi: 10.1007/s00521-018-3849-7.

[24] M. S. Cuthbert and C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data," in Proc. 11th Int. Soc. Music Inf. Retrieval Conf. (ISMIR), Utrecht, The Netherlands, Aug. 2010, pp. 637–642, doi: 10.5281/zenodo.1416114.

[25] C. Harte, M. Sandler, S. Abdallah, and E. Gómez, "Symbolic representation of musical chords: A proposed syntax for text annotations," in Proc. 6th Int. Conf. Music Inf. Retrieval (ISMIR), London, U.K., Sep. 2005, pp. 66–71.

[26] C.-Z. A. Huang et al., "Music Transformer," arXiv preprint arXiv:1809.04281, Dec. 2018, doi: 10.48550/arXiv.1809.04281.

[27] F. Simonetta, F. Carnovalini, N. Orio, and A. Rodà, "Symbolic music similarity through a graph-based representation," in Proc. Audio Mostly 2018 on Sound in Immersion and Emotion (AM '18), New York, NY, USA, 2018, pp. 1–7, doi: 10.1145/3243274.3243301.

[28] D. Temperley, The Cognition of Basic Musical Structures. Cambridge, MA, USA: MIT Press, 2001.

[29] D. L. Brodbeck, "Review of Harmony and Voice Leading, by Edward Aldwell and Carl Schachter," Perspect. New Music, vol. 21, no. 1–2, pp. 425–430, 1982/83.

[30] M. Levine, The Jazz Theory Book. Petaluma, CA, USA: Sher Music, 1995.