The Evolution of Automated Melody Generation

Dr. Janaka V. Wijayakulasooriya

1. Introduction

Automated melody generation has evolved significantly, starting from early stochastic and probabilistic models to powerful Generative AI systems. This lecture note summarizes this progression with key milestones and references.

2. Early Probability and Statistical Techniques

2.1 Markov Chains

- **Concept:** Uses the probability of transitioning from one note to another based on historical data.
- **Application:** Models melody as a sequence where the next note depends only on the current note (first-order) or a few previous notes (higher-order).
- **Example:** Hiller & Isaacson's *Illiac Suite* (1957) – first computer-generated composition using stochastic processes.
- **Limitation:** Lack of global coherence; produces locally consistent but often musically meaningless long-term structures.

Reference: Hiller, L. A., & Isaacson, L. M. (1958). *Experimental Music: Composition with an Electronic Computer*. McGraw-Hill.

2.2 N-gram Models

- **Concept:** Generalizes Markov models to sequences of n items; used in melody continuation and harmonization.

Reference: Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1), 51-73.

2.3 Stochastic Grammars

- **Concept:** Uses probabilistic context-free grammars to model hierarchical structures in music.
- **Example:** Inspired by Lerdahl & Jackendoff's generative theory of tonal music.

Reference: Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.

3. Evolution Towards Machine Learning Techniques

3.1 Hidden Markov Models (HMMs)

- **Application:** Melody transcription, style imitation, improvisation.
- **Strength:** Captures hidden states representing musical intentions or styles.
- **Limitation:** Limited long-range dependency modeling.

Reference: Raphael, C. (2002). A hybrid graphical model for rhythmic parsing. *Artificial Intelligence*, 137(1-2), 217-238.

3.2 Artificial Neural Networks (ANNs)

- Early feed-forward networks mapped input features to pitch sequences but lacked sequential memory.

Reference: Todd, P. M. (1989). A connectionist approach to algorithmic composition. *Computer Music Journal*, 13(4), 27-43.

3.3 Recurrent Neural Networks (RNNs)

- **Application:** Melody generation by learning temporal dependencies.
- **Limitation:** Vanishing gradient problem in long sequences.

Reference: Eck, D., & Schmidhuber, J. (2002). Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. *IEEE NNSP*.

3.4 Long Short-Term Memory Networks (LSTMs)

- **Strength:** Overcame RNN limitations, enabling longer melody sequences with meaningful structures.
- **Example:** Google Magenta's Melody RNN.

Reference: Waite, P. (2016). Generating long-term structure in songs and stories. Magenta Blog. `https://magenta.tensorflow.org/melody-rnn`

4. Generative Adversarial Networks (GANs)

4.1 Basic GAN Application

- **Concept:** Generator and discriminator networks trained adversarially to produce realistic outputs.
- **Application:** Melody generation with stylistic adherence.

Reference: Yang, L. C., Chou, S. Y., & Yang, Y. H. (2017). MidiNet: A convolutional GAN for symbolic-domain music generation. *ISMIR*.

4.2 Conditional GANs

- Conditioning on chord sequences, genres, or rhythmic patterns for controlled generation.

5. Variational Autoencoders (VAEs)

- **Concept:** Encodes melodies into latent spaces for interpolation and smooth transitions between styles.
- **Example:** MusicVAE by Google Magenta.

Reference: Roberts, A., et al. (2018). Hierarchical latent vector model. *ICML.*

6. Transformer Models

6.1 Basic Transformer

- **Application:** Uses self-attention to capture long-range dependencies better than RNNs.
- **Example:** Music Transformer by Huang et al.

Reference: Huang, C. A., et al. (2018). Music Transformer. *ICLR.*

6.2 Advantages

- Better scalability to long sequences.
- Superior coherence in melody and accompaniment generation.

7. Recent Advances: Generative AI Systems

7.1 Suno AI

- **Description:** Generates full songs with lyrics, vocals, and accompaniment from text prompts using large transformer models.

Reference: Suno AI. (2024). `https://suno.ai/`

7.2 Google Musix (MusicLM)

- **Description:** Generates high-fidelity music from textual descriptions using hierarchical sequence-to-sequence modeling.

Reference: Agostinelli, A., et al. (2023). MusicLM: Generating Music From Text. *arXiv:2301.11325.*

7.3 Other Notable Systems

- **OpenAI Jukebox:** End-to-end neural net for raw audio music generation with lyrics.

  Reference: Dhariwal, P., et al. (2020). Jukebox. OpenAI. `https://openai.com/research/jukebox`

- **Riffusion:** Stable Diffusion adapted for spectrogram generation converted back to audio.

  Reference: Forsgren, H., & Martiros, S. (2022). Riffusion. *arXiv:2212.07650.*

8. Summary of Evolutionary Milestones

Technique	**Key Feature**	**Limitation**
Markov Models	Local transition probabilities	No long-term coherence
HMMs	Hidden state modeling	Limited to short-term dependencies
RNNs	Temporal sequence learning	Vanishing gradients
LSTMs	Long sequence memory	Limited style generalization
GANs	Realistic data generation	Training instability
VAEs	Latent interpolation	Less sample realism
Transformers	Long-range attention	High computational cost
GenAI Models	End-to-end controlled music generation	Requires vast data & compute

9. Conclusion

The field has progressed from simple probability models to transformer-based generative AI systems capable of producing human-like compositions conditioned on textual, stylistic, and emotional prompts, redefining music creativity and human-computer co-creation paradigms.