Inference-Time Control of Tonal Tension in Symbolic Music Generation

Maral Ebrahimzadeh, Gilberto Bernardes, Sebastian Stober

Otto-von-Guericke-University Magdeburg, Germany – Artificial Intelligence Lab

contact: maral.ebrahimzadeh@ovgu.de

Motivation

- Symbolic music generation has advanced with LLMs and Transformers.
- Explicit control over high-level features remains limited.
- Tonal tension, a key compositional feature, is still underexplored.
- The Tonal Interval Vectors (TIV) framework offers an efficient, perceptually grounded way to compute tension.
- Training-time control requires retraining, while inference-time control is flexible.
- Need a practical method combining local quality (probability + diversity) with global tension shaping (target curve alignment).

TIV based Tonal Tension Model

- Distance between current chord and previous chord

$$d_1(T_i, T_{i-1}) = \mu(T_{i-1}, T_i)$$

- Distance between current chord and key

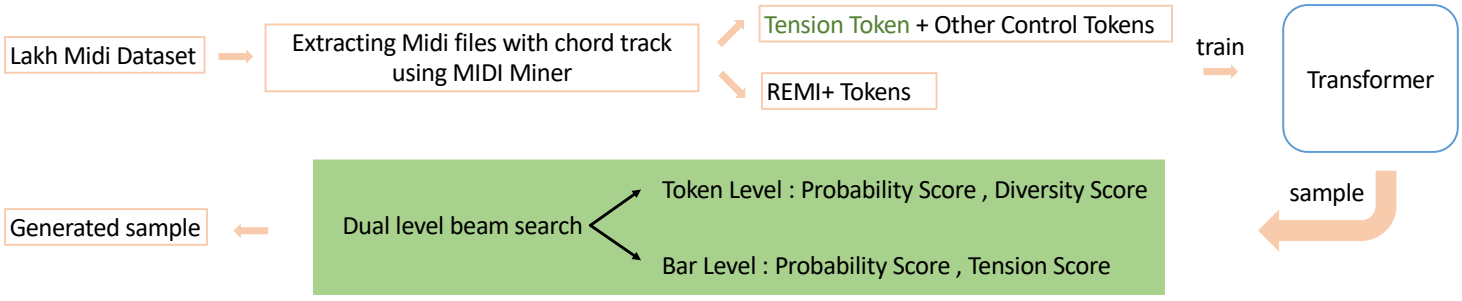
$$d_2(T_i, T_{key}) = \theta(T_i, T_{key})$$

- Distance between current chord and tonal function

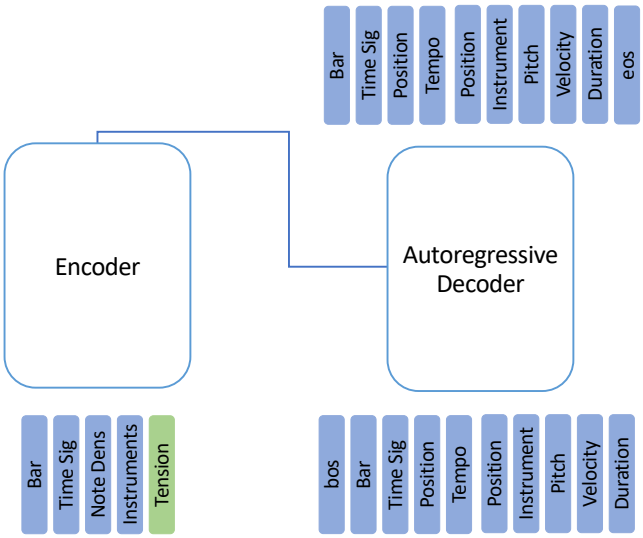
$$d_3(T_i - T_{key}, T_f) = \theta(T_i - T_{key}, T_f)$$

- Dissonance $1 - \frac{\|T_i\|}{\|T_{max}\|}$
- Voice Leading $m(T_i, p) = \sum_{l=1}^{V} \frac{1}{e^{0.05 s \mu(T_{n_{l_i}}, T_{n_{l_{i-1}}})}}$

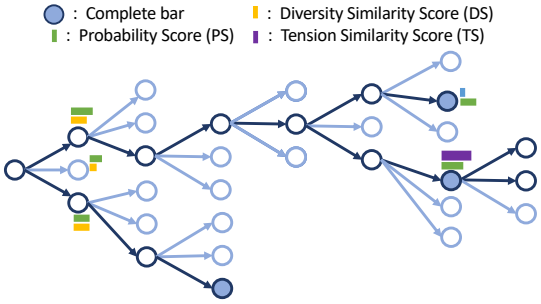
Method Overview



Model and Input Representation



Inference: Dual Level Beam Search



Token Level:

$$DS = \text{duration diversity} + \text{pitch diversity} + 3\text{ gram pitch entropy}$$

$$\text{token level score} = PS + weight_{DS} \times DS$$

Bar Level:

$$TS = \begin{cases} Correalation & if\ Var > 0.0001 \\ Absolute\ Difference & otherwise \end{cases}$$

$$\text{bar level score} = PS + weight_{TS} \times TS$$

Result

Model	Inference	Instrument F1	Note Density	Groove Similarity	Tension Correlation
Baseline	**Normal**	0.82	0.88	0.52	0.16
Baseline + tension	**Normal**	0.83	0.62	0.54	0.18
Baseline + tension	**Dual Beam**	0.86	0.85	0.56	0.50

Tension Curve