BEAT TRACKING FOR REAL-TIME HUMAN-COMPUTER LIVE IMPROVISATION

Joris MONNET[1], **Michael WOLLNY**[2], and **Martin ROHRMEIER**[1]

[1]**Digital and Cognitive Musicology Lab (EPFL)**, Lausanne, Switzerland
[2]**Hochschule für Musik und Theater Felix Mendelssohn Bartholdy**, Leipzig, Germany

ABSTRACT

Beat tracking is a fundamental task in computational musicology, with applications ranging from music analysis to interactive performance systems. While many existing approaches achieve high accuracy in offline settings, they typically rely on access to the complete musical signal and are not designed for real-time use. In live performance and human–computer interaction scenarios, beat tracking must operate under strict latency constraints while remaining robust to expressive timing, tempo fluctuations, and structural variability. In this paper, we evaluate a set of beat tracking methods within a real-time framework, with a particular focus on MIDI-based input. We compare several classes of approaches, including inter-onset interval models, onset-driven methods, and frequency-domain techniques, within a modular streaming pipeline. Experiments are conducted on classical and jazz piano datasets, including augmented data to simulate tempo variation. Our results show that while most methods provide reliable tempo estimation, they struggle to maintain precise and stable beat localization under real-time constraints. This discrepancy highlights a limitation of current evaluation paradigms, which may overestimate performance by not fully accounting for temporal alignment in interactive settings. These findings underline the need for beat tracking approaches and evaluation frameworks specifically designed for real-time musical interaction.

Keywords: Beat tracking, Beat induction, real-time music analysis, MIDI, interactive music systems, tempo estimation

1. INTRODUCTION

Beat tracking is a long-standing and fundamental problem in computational musicology, concerned with identifying the temporal pulse that underlies musical structure. It plays a central role across a wide range of applications, including music analysis, large-scale corpus studies, interactive music software, and performance systems. Accurately estimating the beat enables higher-level reasoning about rhythm, synchronization, and musical form.

Beat refers to the basic unit of time in music, representing a typically regular pulse and forming the foundation of temporal organization in performance and perception [1]. Despite its apparent simplicity, beat perception is highly complex [2] and influenced by cognitive factors such as culture [3, 4] and age [5, 6], as well as underlying perceptual mechanisms [7–10]. Beats are organized hierarchically through meter (e.g., 4/4, 3/4, 6/8), which structures strong and weak beats.

Beat induction refers to identifying the underlying pulse of a musical signal, including both *tempo estimation* and *beat inference*. Notes aligned with the beat are generally more perceptually salient [11], and different approaches target different rhythmic levels, including meter [12], tatum [13, 14], or several metrical layers at once [15, 16], while others focus primarily on tempo estimation [17–21].

Most existing beat tracking approaches assume offline access to the complete musical signal, enabling the use of future context to refine predictions. However, in interactive performance and Human–Computer Interaction (HCI) scenarios, systems must operate incrementally under strict timing constraints, producing reliable beat estimates with minimal latency while adapting to expressive variability such as tempo fluctuations, microtiming deviations, and structural changes.

This work is motivated by the development of a human–machine collaborative system for live, autonomous improvisation called *LARS*, which requires robust beat tracking in real time. It operates in the MIDI [22] domain, providing direct access to structured musical events (timing, velocity, pitch), enabling precise and computationally efficient processing [23]. However, existing methods do not provide sufficiently reliable performance in this context, particularly in the presence of dynamic tempo changes, shifting time signatures, and complex rhythmic structures like polyrythms. To address this limitation, we evaluate and adapt a set of beat-tracking methods within a real-time framework, with a particular focus on their robustness, responsiveness, and ability to maintain synchronization in live musical interaction.

2. RELATED WORK

Beat induction has been extensively studied, with most approaches focusing on offline analysis of audio signals. These methods typically rely on feature extraction techniques to identify rhythmic patterns, using cues such as onsets, tempo, and harmonic changes [19, 24, 25]. Onset detection, often based on spectral flux, is commonly used as a first step, followed by alignment techniques such as

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

dynamic programming to estimate beat positions [26].

Real-time beat tracking, often formulated as *process models*, was introduced by Scheirer [19] to estimate and track beats incrementally without access to the full signal. Early approaches include Scheirer's multi-band analysis [19] and Dixon's adaptive synchronization method [24], which explicitly address tempo variability. Subsequent work introduced probabilistic frameworks such as Hidden Markov Models to improve robustness [26, 27]. More recent approaches leverage deep learning techniques to enhance performance in complex scenarios [15, 27–31]. Recurrent neural networks have been used for beat detection [28], and hybrid systems combine neural models with real-time adaptation mechanisms [18, 32]. However, these approaches are primarily designed for audio signals and often require substantial computational resources, which can limit their applicability in real-time interactive settings.

Despite these advances, real-time beat tracking remains challenging. Without access to global context, process models must continuously evaluate and correct predictions, making them more prone to errors [24]. They must also handle noisy inputs, extreme tempo variations, and octave errors [33, 34], where different metrical levels (e.g., tatum vs. beat) are confused. These challenges are amplified in complex musical contexts such as jazz and free improvisation [35, 36], as well as in non-Western musical traditions [37–39], where syncopation, polyrhythms, and irregular meters complicate beat induction.

Overall, while prior work provides strong foundations for beat tracking, it remains largely centered on offline and audio-based scenarios, leaving open challenges for real-time, MIDI-based systems in interactive musical contexts.

3. DESIGN

3.1 Naive model: Inter-Onset Intervals based Beat Detection

Beat detection can be approached through Inter-Onset Intervals (IOIs), i.e. temporal distances between note onsets [24]. Recurrent IOI patterns reveal the underlying rhythmic structure, including beat and bar periodicities. Two IOI formulations are considered: all-pairs (*M to M*) and forward differences (*M to N*):

$$IOI_{M\to M} = \left\{ |t_n - t_m| \;\middle|\; \forall n, m \in \{1, \ldots, N\} \right\} \quad (1)$$

$$IOI_{M\to N} = \left\{ t_j - t_i \;\middle|\; i < j \right\} \quad (2)$$

where t_n denotes the onset time of the n-th note, N is the total number of note onsets, and i, j, m, and n are onset indices. Extensions incorporate duration (offsets) and velocity, assuming stronger beats align with louder notes. While computationally efficient, IOI methods struggle with complex rhythms (see Section 5).

3.2 Onset-Based Beat Detection

This model assumes onsets are more likely to occur on beats. Onsets are represented as a discrete spike train [40]:

$$s[n] = \begin{cases} 1 & \text{if there is an onset at frame } n \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Normalized autocorrelation models beat periodicity:

$$R[\tau] = \frac{\sum_n s[n]\, s[n+\tau]}{\sum_n s[n]^2} \quad (4)$$

where $R[\tau]$ is the autocorrelation at lag τ, and τ is the time lag measured in frames.

The first significant peak gives the beat period T (in frames), and tempo (in Beats Per Minute) is computed as:

$$\text{Tempo} = \frac{60 \cdot f_s}{T \cdot H} \quad (5)$$

where f_s is the sampling rate, H the hop size (in samples), and T the estimated beat period in frames. This approach is efficient and suitable for real-time use but degrades under syncopation or sparse textures.

3.3 Fast Fourier Transform (FFT) and Autocorrelation

FFT-based methods identify periodicities in the frequency domain [40, 41] using the discrete Fourier transform:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N} \quad (6)$$

where $X[k]$ is the Fourier coefficient at frequency bin k, $x[n]$ is the input signal at sample n, and N is the number of samples in the analysis window.

Autocorrelation reinforces periodicity detection:

$$R[\tau] = \sum_n x[n]\, x[n+\tau] \quad (7)$$

where $R[\tau]$ denotes the autocorrelation at lag τ.

Shannon entropy,

$$H = -\sum_k P(k) \log P(k), \quad (8)$$

measures spectral regularity, where $P(k)$ is the normalized spectral probability at frequency bin k, and H denotes the Shannon entropy. Low entropy indicates strong periodicity. These features are combined to estimate beat structure [41].

Dynamic programming further refines beat alignment [26]. A reference period

$$T_{\text{ref}} = \frac{1}{M-1} \sum_{i=2}^{M} (t_i - t_{i-1}) \quad (9)$$

guides alignment via a penalty:

$$P(\tau) = -\alpha \left(\log_2 \frac{\tau}{T_{\text{ref}}} \right)^2 \quad (10)$$

where T_{ref} is the reference beat period, M is the number of detected onset times, t_i is the onset time of the i-th note,

$P(\tau)$ is the penalty associated with a candidate interval τ, and α controls the penalty strength.

The optimal beat sequence is obtained recursively:

$$A(n) = N(n) + \max_{m<n}[A(m) + P(n-m)] \tag{11}$$

where $A(n)$ is the accumulated score at frame n, $N(n)$ is an onset strength (or novelty) function, and $P(n-m)$ is the transition penalty between candidate beat positions m and n. Backtracking yields beat times B, from which period and phase are derived. This approach improves robustness to tempo variations.

3.4 The Predominant Local Pulse (PLP) method

The PLP method [42, 43] estimates local periodicity using windowed sinusoidal kernels. Onsets are encoded as a spike train:

$$e[n] = \begin{cases} 1 & \text{if there is an onset at frame } n \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Local periodicity is computed for different tempo candidates:

$$p_f[n] = \sum_{\tau} e[n+\tau] \sin(2\pi f \tau) \tag{13}$$

where $p_f[n]$ is the local periodicity estimate at frame n for tempo candidate f, τ is the lag variable, and f denotes the candidate pulse frequency.

These are aggregated across frequencies:

$$\text{PLP}[n] = \sum_{f} w(f)\, p_f[n] \tag{14}$$

where $w(f)$ is the weighting applied to tempo candidate f.

Autocorrelation of PLP$[n]$ yields the dominant period τ_{peak}, from which tempo is derived:

$$\text{Tempo} = \frac{60 \cdot f_s}{\tau_{\text{peak}} \cdot H} \tag{15}$$

where τ_{peak} is the lag corresponding to the dominant autocorrelation peak.

PLP is effective for time-varying tempo but less reliable for precise beat localization.

Neural approaches improve robustness in complex scenarios [18, 28, 32, 44], but remain limited for real-time use due to computational cost and data requirements. This is particularly problematic for niche genres such as free jazz, where datasets are scarce and overfitting is likely. For these reasons, we do not use neural models. However, emerging lightweight and on-device approaches [45, 46] may make them viable for future real-time systems.

4. IMPLEMENTATION

4.1 Datasets

Two MIDI piano datasets were used, consistent with our own system real-time input constraints. The *ASAP dataset* [47, 48] contains 236 scores and 1066 performances of Western classical piano music. Each MIDI file includes aligned beat annotations used as ground truth, derived via automated methods described in [48, 49]. The dataset includes both metronomic and expressive performances, introducing tempo variability. A second dataset was created with pianist Michael Wollny, consisting of 14 short jazz pieces with controlled tempo initialization. Ground truth is derived by estimating tempo from several repeated notes at the beginning of each sample, then extrapolating automatically beat positions with tolerance adjustments [50].

To increase diversity, MIDI files were augmented using linear tempo changes (50%, 66%, 150%, 200%), including acceleration–deceleration profiles, resulting in 126 total pieces. This better reflects expressive performance compared to uniform time-stretching [51].

4.2 Evaluation Methods

Evaluation follows established metrics [50], complemented by two evaluation methods introduced in this work: (i) a visual Single Beat Comparison for qualitative analysis of timing errors, and (ii) the Mean Tempo Estimation Error (μTEE), a complementary quantitative measure of tempo estimation accuracy.

F-measure/Precision/Recall are used to evaluate predicted beats matching to annotations within a tolerance window (e.g., 10–20% of beat duration [34, 50, 52] or $\pm$70 ms [44, 53]).

P-score [50] provides a softer alternative to the F-measure by assigning partial credit to beats that are close to the ground truth instead of treating predictions as simply correct or incorrect. Small timing errors receive higher weights, whereas predictions outside the tolerance window contribute nothing. This makes the metric more sensitive to gradual timing deviations than binary matching measures.

$$P = \frac{1}{N} \sum_{n=1}^{N} \max\left(0, 1 - \frac{|\Delta t_n|}{\tau}\right) \tag{16}$$

where N is the total number of annotated beats, Δt_n is the timing error between the n-th predicted and annotated beat, and τ is the tolerance window.

Continuity metrics (CMLc, CMLt, AMLc, AMLt) [54] evaluate whether correct beat tracking is maintained over extended portions of a performance rather than on isolated beats. CML metrics require the correct metrical level to be maintained, whereas AML metrics also accept musically related metrical interpretations such as double or half tempo. These metrics are particularly useful for distinguishing algorithms that occasionally recover from errors from those that maintain a stable beat estimate throughout a piece.

Information Gain measures reduction in rhythmic uncertainty [55, 56]:

$$IG = H_{\text{baseline}} - H_{\text{model}} \tag{17}$$

where H_{baseline} is the entropy of the baseline beat distribution, and H_{model} is the entropy of the beat distribution

produced by the evaluated model. Higher values indicate that beat predictions are concentrated around the annotated beat positions rather than being randomly distributed. Unlike threshold-based metrics, IG evaluates the overall distribution of timing errors.

Cemgil's metric [57] uses a Gaussian error weighting:

$$W(x) = \exp\left(-\frac{x^2}{2\sigma_e^2}\right) \tag{18}$$

where x is the beat timing error, and σ_e is the standard deviation controlling the tolerance of the weighting.

$$Cem_{acc} = \frac{\sum_j \max_b W(\gamma_b - a_j)}{\frac{B+J}{2}} \tag{19}$$

where B is the number of predicted beats, J is the number of annotated beats, γ_b is the time of the b-th predicted beat, and a_j is the time of the j-th ground truth annotation.

Goto's metric [58] evaluates whether an algorithm produces a long, continuous sequence of consistently accurate beat estimates. Rather than averaging individual beat errors, it requires several consistency conditions to be satisfied simultaneously, making it a strict measure of reliable beat tracking over time. Consequently, a single local improvement is insufficient to obtain a high score.

The normalized beat error, ζ_j, is defined as:

$$\zeta_j = \begin{cases} \frac{|\gamma_b - a_j|}{\Delta_j/2} & \text{if } a_{j-1} + \frac{\Delta_{j-1}}{2} \leq \gamma_b < a_j + \frac{\Delta_j}{2}, \\ 1 & \text{otherwise} \end{cases} \tag{20}$$

where:

- γ_b is the time of the b-th predicted beat,
- a_j is the time of the j-th ground truth annotation,
- $\Delta_j = a_j - a_{j-1}$ is the inter-annotation interval.

A subsequence ζ_k of valid beat errors ($\zeta_j < 0.35$) is extracted, and three conditions are checked:

$$I_1(\zeta_k) = \begin{cases} 1 & \text{if } \frac{k_s}{k_e} < 0.75 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

$$I_2(\zeta_k) = \begin{cases} 1 & \text{if } \text{mean}(\zeta_k) < 0.2 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

$$I_3(\zeta_k) = \begin{cases} 1 & \text{if } \text{std}(\zeta_k) < 0.2 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where I_1, I_2, and I_3 are indicator functions evaluating the three Goto criteria, and k_s and k_e are the start and end indices of the subsequence ζ_k.

The overall metric is defined as:

$$Goto_{acc} = I_1(\zeta_k)\, I_2(\zeta_k)\, I_3(\zeta_k) \times 100\% \tag{24}$$

Single Beat Comparison (proposed in this work) provides a visual representation of the distribution of timing errors relative to each annotated beat. Unlike aggregate metrics, it reveals systematic biases, such as predictions consistently occurring slightly before or after the beat, as well as error caused by metrical ambiguities. It complements the quantitative metrics by helping identify characteristic failure modes of different algorithms.

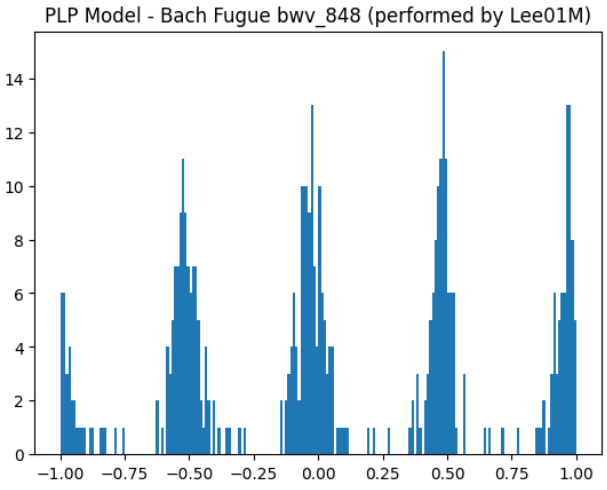


Figure 1. Single Beat Comparison (Poor Performance)

The x-axis in Figures 1 represents the total range spanning from the half-beat before the target beat (-1) to the half-beat after it (+1), with the target beat itself positioned at 0. These half-beats are not fixed in absolute time but are always defined as 50% of the ground truth beat duration, either leading up to the target beat or following it. For instance, Figure 1 illustrates an innacurate model where a greater number of beats are detected slightly later, aligning more often with the quarter-beat position after the target beat, Which is a sixteenth note delay in this case (4/4 meter).

Performance is assessed via peak distribution and standard deviation, capturing both accuracy and consistency. This representation highlights systematic offsets (e.g., half- or quarter-beat errors) and enables direct comparison between models.

Mean Tempo estimation error (µTEE)

The Mean Tempo Estimation Error (µTEE) is proposed in this work as a complementary evaluation metric. It measures the accuracy of a model in estimating the overall tempo of a piece, independently of individual beat alignment. It is defined as the absolute difference between the mean interval of the detected beats and the mean interval of the ground truth beat times:

$$\text{µTEE} = |\mu(\text{beats}) - \mu\left(\Delta(\text{ground_truth})\right)|$$

where:

- $\mu(\cdot)$ denotes the arithmetic mean.
- $\mu(\text{beats})$ represents the mean interval between consecutive detected beats.
- $\Delta(\text{ground_truth})$ represents the consecutive time differences in the ground truth beat times (e.g., $\Delta(\text{ground_truth}) = \text{np.diff}(\text{ground_truth})$).

Table 1. Main performance metrics of beat tracking models for 3s and 5s sliding windows. Only the most relevant metrics are shown: F1, P-Score, Cemgil, and Goto (%).

Model	SW	F1	P	Cemgil	Goto %
Naive IOI	3	0.24	0.36	0.24	93.84
Naive IOI	5	0.20	0.34	0.21	65.41
IOI	3	0.24	0.35	0.24	92.72
IOI	5	0.19	0.35	0.19	65.69
IOI/Dur	3	0.18	0.28	0.20	5.18
IOI/Dur	5	0.16	0.31	0.18	8.82
IOI/Vel	3	0.25	**0.38**	0.26	94.82
IOI/Vel	5	0.21	0.36	0.21	79.83
IOI/Vel/Dur	3	0.17	0.31	0.19	9.38
IOI/Vel/Dur	5	0.17	0.30	0.19	6.51
Onset	3	0.17	0.30	0.17	97.97
Onset	5	0.16	0.30	0.16	96.99
Oscillator	3	0.28	0.34	**0.29**	9.66
Oscillator	5	**0.30**	**0.37**	**0.31**	10.49
Ellis	3	**0.29**	0.35	**0.29**	97.82
Ellis	5	**0.29**	0.34	**0.29**	**98.21**
PLP	3	0.05	0.33	0.07	4.06
PLP	5	0.04	0.28	0.05	8.96

Table 2. Secondary metrics: strict and flexible beat consistency with mean tempo error.

Model	SW	CMLc %	AMLc %	µTEE (s)
Naive IOI	3	2.54	6.33	0.107
Naive IOI	5	3.26	6.61	0.102
IOI	3	2.69	4.83	**0.087**
IOI	5	4.72	9.51	0.109
IOI/Dur	3	2.09	4.24	0.094
IOI/Dur	5	3.35	6.65	0.111
IOI/Vel	3	4.30	9.26	0.107
IOI/Vel	5	3.51	6.71	0.109
IOI/Vel/Dur	3	2.98	6.83	0.108
IOI/Vel/Dur	5	2.36	4.97	0.100
Onset	3	2.11	5.09	0.102
Onset	5	2.25	5.27	0.100
Oscillator	3	2.89	6.20	**0.083**
Oscillator	5	5.36	7.97	**0.083**
Ellis	3	3.76	7.24	0.125
Ellis	5	3.53	6.75	0.151
PLP	3	**7.23**	**14.40**	0.259
PLP	5	**8.66**	**15.42**	0.568

- $\mu(\Delta(\text{ground_truth}))$ is the mean of these time differences, corresponding to the average ground truth beat interval.

The different evaluation metrics demonstrate complementary properties of beat tracking. For example, an algorithm may achieve a high F-measure by consistently predicting beats within the tolerance window, while still receiving a lower Cemgil score due to small but systematic timing offsets. Similarly, µTEE may indicate accurate tempo estimation even when beat localization metrics reveal poor synchronization.

5. RESULTS AND DISCUSSION

Tables 1 and 2 display the results. Each model is represented by two rows corresponding to 3s and 5s sliding windows. Metrics requiring a margin use ±50ms around the ground truth. Longer sliding windows (5s) tend to lower F1 and P-scores, suggesting that a 3s window provides a better balance between temporal accuracy and overall performance. Additionally, one of the evaluated metrics was the computation time for each model, and unsurprisingly, the 5s window requires more time to process than the 3s window.

Beat tracking involves two components: tempo estimation and beat localization. Mean tempo error (µTEE) shows that several models estimate tempo accurately ($<$100ms), so low F1 scores are mainly due to imprecise beat placement.

The PLP model captures overall rhythmic structure (high AMLc/Info Gain) but fails at precise beat placement (low F1, Cemgil, and Goto), making it less suitable for practical beat tracking. Ellis and Oscillator provide better alignment, consistency, and robustness.

IOI-based models like Naive IOI and IOI+Vel achieve high Goto scores, showing robustness on simpler rhythmic patterns, while PLP struggles on complex excerpts. Ellis and Oscillator show consistently high F1 and Cemgil scores, highlighting practical reliability.

On datasets with expressive performances (e.g., Wollny and ASAP), Ellis, Oscillator, and IOI/Vel models perform better than the others, indicating their suitability for interactive and real-time contexts.

6. CONCLUSION

This paper presented a systematic evaluation of beat tracking methods in a real-time performance context, with a particular focus on MIDI-based input and interactive musical scenarios. Although many beat tracking approaches have been developed and evaluated for offline or audio-based applications [50, 55], their suitability for real-time symbolic interaction has received comparatively little attention. Our evaluation shows that, under these constraints, achieving reliable beat tracking remains challenging, particularly when low latency, adaptability to tempo fluctuations, and robustness to expressive timing are required.

Across all evaluated models, a clear distinction emerges between tempo estimation and beat localization. Consistent with previous observations [57, 58], most methods are able to estimate tempo with reasonable accuracy, even under moderate tempo variation, but fail to maintain precise and stable beat alignment over time. This limitation becomes particularly evident in expressive performances, such as jazz improvisation or tempo-modulated excerpts, where microtiming deviations and structural changes challenge the assumptions of periodicity underlying many beat tracking systems.

Among the tested approaches, oscillator-based and dynamic programming methods achieve the most balanced

performance, combining relatively strong alignment accuracy with adaptability to tempo changes. This aligns with prior work emphasizing the importance of continuous temporal adaptation [26,59]. In contrast, methods such as Predominant Local Pulse (PLP) [26] capture global rhythmic structure but lack the temporal precision required for reliable synchronization, while IOI-based approaches remain sensitive to rhythmic complexity despite their efficiency in symbolic domains.

These results highlight a fundamental gap between existing evaluation paradigms and the requirements of real-time musical interaction. Standard metrics such as F-measure, Cemgil accuracy, or Information Gain [55, 57] may overestimate performance by conflating tempo accuracy with precise beat alignment, thereby masking limitations that are critical in interactive contexts.

More broadly, this study reinforces the need to reconsider beat tracking from the perspective of process rather than offline prediction, particularly in MIDI-based systems where structured temporal information is directly available. Despite this advantage, MIDI-based real-time beat tracking remains underexplored, especially in comparison to audio-driven approaches.

For applications such as human–machine improvisation, reliable beat tracking remains an open challenge. Future work should focus on hybrid approaches that combine local responsiveness with longer-term temporal coherence, as well as evaluation frameworks specifically designed for real-time contexts, where continuity, latency, and interaction play a central role.

Acknowledgments

This project was in part funded by the Swiss National Science Foundation through the AGORA grant (no.224008). Aspects of the project were also funded in the context of the SNF grants no. 182811 and 215701. Martin Rohrmeier and his lab at EPFL have been funded through Mr Claude Latour.

7. REFERENCES

[1] J. London, *Hearing In Time: Psychological Aspects of Musical Meter*. Oxford University Press, Sep. 2004.

[2] P. Grosche, M. Müller, and C. S. Sapp, "What Makes Beat Tracking Difficult? A Case Study on Chopin Mazurkas." in *Proceedings of the 11th International Society for Music Information Retrieval Conference*. Utrecht, Netherlands: ISMIR, Aug. 2010, pp. 649–654.

[3] D. J. Cameron, J. Bentley, and J. A. Grahn, "Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping," *Frontiers in Psychology*, vol. 6, p. 366, Apr. 2015.

[4] N. Jacoby, R. Polak, J. A. Grahn, D. J. Cameron, K. M. Lee, R. Godoy, E. A. Undurraga, T. Huanca, T. Thalwitzer, N. Doumbia, D. Goldberg, E. H. Margulis, P. C. M. Wong, L. Jure, M. Rocamora, S. Fujii, P. E. Savage, J. Ajimi, R. Konno, S. Oishi, K. Jakubowski, A. Holzapfel, E. Mungan, E. Kaya, P. Rao, M. A. Rohit, S. Alladi, B. Tarr, M. Anglada-Tort, P. M. C. Harrison, M. J. McPherson, S. Dolan, A. Durango, and J. H. McDermott, "Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries," *Nature Human Behaviour*, vol. 8, no. 5, pp. 846–877, May 2024.

[5] E. C. Thompson, T. White-Schwoch, A. Tierney, and N. Kraus, "Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience," *PLoS ONE*, vol. 10, no. 6, p. e0128839, Jun. 2015.

[6] H. Honing, *Music Cognition: The Basics*. London: Routledge, Sep. 2021.

[7] ——, "Without it no music: beat induction as a fundamental musical trait," *Annals of the New York Academy of Sciences*, vol. 1252, pp. 85–91, Apr. 2012.

[8] H. Merchant, J. Grahn, L. Trainor, M. Rohrmeier, and W. Fitch, "Finding the beat: A neural perspective across humans and non-human primates," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 370, Mar. 2015.

[9] E. W. Large and J. S. Snyder, "Pulse and Meter as Neural Resonance," *Annals of the New York Academy of Sciences*, vol. 1169, no. 1, pp. 46–57, 2009.

[10] N. P. McAngus Todd, D. J. O'Boyle, and C. S. Lee, "A Sensory-Motor Theory of Rhythm, Time Perception and Beat Induction," *Journal of New Music Research*, vol. 28, no. 1, pp. 5–28, Mar. 1999.

[11] A. Danielsen, K. Nymoen, E. Anderson, G. Câmara, M. Langerød, M. Thompson, and J. London, "Where Is the Beat in That Note? Effects of Attack, Duration, and Frequency on the Perceived Timing of Musical and Quasi-Musical Sounds," *Journal of experimental psychology. Human perception and performance*, vol. 45, pp. 402–418, Mar. 2019.

[12] A. McLeod and M. Steedman, "Meter Detection and Alignment of MIDI Performance," *Proceedings of the 19th International Society for Music Information Retrieval Conference*, pp. 113–119, Sep. 2018.

[13] J. Seppanen, "Tatum grid analysis of musical signals," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, Oct. 2001, pp. 131–134.

[14] J. A. Bilmes, "Timing is of the essence : perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm," Thesis, Massachusetts Institute of Technology, 1993.

[15] M. Heydari, F. Cwitkowitz, and Z. Duan, "BeatNet: CRNN and Particle Filtering for Online Joint Beat

Downbeat and Meter Tracking," in *22th International Society for Music Information Retrieval Conference, ISMIR*, 2021.

[16] T. Oyama, R. Ishizuka, and K. Yoshii, "Phase-Aware Joint Beat and Downbeat Estimation Based on Periodicity of Metrical Structure," *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, pp. 493–499, Nov. 2021, conference Name: International Society for Music Information Retrieval Conference (ISMIR 2021). [Online]. Available: https://zenodo.org/records/5624517

[17] A. Gkiokas, V. Katsouros, G. Carayannis, and T. Stajylakis, "Music tempo estimation and beat tracking by applying source separation and metrical relations," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 421–424.

[18] H. Schreiber and M. Müller, "A Single-step Approach to Musical Tempo Estimation using a Convolutional Neural Network," *Proceedings of the 19th International Society for Music Information Retrieval Conference*, pp. 98–105, Sep. 2018.

[19] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *The Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, Jan. 1998.

[20] G. Peeters, "Template-Based Estimation of Time-Varying Tempo," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 067215, Dec. 2006.

[21] M. F. McKinney and D. Moelants, "Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels?" *Music Perception: An Interdisciplinary Journal*, vol. 24, no. 2, pp. 155–166, 2006.

[22] J. Rothstein, *MIDI: A comprehensive introduction*. AR Editions, Inc., 1995, vol. 7.

[23] M. Goto, "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds," *Journal of New Music Research*, vol. 30, Sep. 2002.

[24] S. Dixon, "Automatic Extraction of Tempo and Beat From Expressive Performances," *Journal of New Music Research*, vol. 30, no. 1, pp. 39–58, Mar. 2001.

[25] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano, "An experimental comparison of audio tempo induction algorithms," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1832–1844, Sep. 2006.

[26] D. P. W. Ellis, "Beat Tracking by Dynamic Programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, Mar. 2007.

[27] F. Krebs, S. Böck, and G. Widmer, "An Efficient State-Space Model for Joint Tempo and Meter Tracking," in *International Society for Music Information Retrieval Conference*, 2015.

[28] S. Böck and M. Schedl, "Enhanced Beat Tracking with Context-Aware Neural Networks," in *Proceedings of the 14th International Conference on Digital Audio Effects, DAFx 2011*, Sep. 2011.

[29] Y.-C. Chuang and L. Su, "Beat and Downbeat Tracking of Symbolic Music Data Using Deep Recurrent Neural Networks," *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2020.

[30] M. Heydari and Z. Duan, "Don't look back: An online beat tracking method using RNN and enhanced particle filtering," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 236–240.

[31] F. Krebs, S. Böck, M. Dorfer, and G. Widmer, "Downbeat Tracking Using Beat Synchronous Features with Recurrent Neural Networks," in *International Society for Music Information Retrieval Conference*, 2016.

[32] S. Böck, F. Krebs, and G. Widmer, "Joint Beat and Downbeat Tracking with Recurrent Neural Networks," in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, M. I. Mandel, J. Devaney, D. Turnbull, and G. Tzanetakis, Eds., Aug. 2016, pp. 255–261.

[33] A. Elowsson and A. Friberg, "Modeling the perception of tempo," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3163–3177, Jun. 2015.

[34] L. M. Smith, "Beat Critic: Beat Tracking Octave Error Identification By Metrical Profile Analysis," *International Society for Music Information Retrieval Conference*, 2010.

[35] F. Benadon, "Gridless Beats," *Perspectives of New Music*, vol. 47, no. 1, pp. 135–164, 2009.

[36] M. Pfleiderer and K. Frieler, "The Jazzomat Project: Issues and Methods for the Automatic Analysis of Jazz Improvisations," in *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, R. Bader, C. Neuhaus, and U. Morgenstern, Eds. Frankfurt am Main and Bern: Peter Lang, 2010, pp. 279–295.

[37] D. P. Wulandari, A. Tjahyanto, and Y. K. Suprapto, "Gamelan Music Onset Detection based on Spectral Features," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 11, no. 1, pp. 107–118, Mar. 2013.

[38] P. Toiviainen and T. Eerola, "Where is the beat?: Comparison of Finnish and South African listeners," in *Proceedings of the 5th Triennial ESCOM Conference*. Hanover University of Music and Drama, 2003, pp. 501–504.

[39] V. Iyer, M. Wright, D. Wessel, and J. Bilmes, "A novel representation for rhythmic structure," in *Proceddings International Computer Music Conference*

1997. Thessaloniki, Greece: The International Computer Music Association, 1997, pp. 97–100, section: International Computer Music Conference 1997. Proceedings: Thessaloniki, Hellas. 25-30 september 1997.

[40] M. Müller, *Fundamentals of Music Processing*. Springer, Jan. 2015.

[41] D. Eck and N. Casagrande, "Finding Meter in Music Using An Autocorrelation Phase Matrix and Shannon Entropy." in *Proceedings of the 6th International Conference on Music Information Retrieval*, London, UK, Sep. 2005, pp. 504–509.

[42] P. Grosche and M. Muller, "Extracting Predominant Local Pulse Information From Music Recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1688–1701, Aug. 2011.

[43] C.-Y. Chiu, M. Müller, M. Davies, A. Su, and y.-h. Yang, "Local Periodicity-Based Beat Tracking for Expressive Classical Piano Music," Aug. 2023.

[44] S. Murgul and M. Heizmann, "Beat and Downbeat Tracking in Performance Midi Using an End-to-End Transformer Architecture," in *Proceedings of the 22nd Sound and Music Computing Conference (SMC2025), Graz, July 2025*. Graz, Austria: Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Jul. 2025, pp. 183–189, conference Name: Sound and Music Computing 2025 (SMC 2025) ISBN: 9783200106420.

[45] F. Chen, S. Li, J. Han, F. Ren, and Z. Yang, "Review of Lightweight Deep Convolutional Neural Networks," *Archives of Computational Methods in Engineering*, vol. 31, no. 4, pp. 1915–1937, May 2024.

[46] K. Saravanan and A. Z. Kouzani, "Advancements in On-Device Deep Neural Networks," *Information*, vol. 14, no. 8, p. 470, Aug. 2023.

[47] F. Foscarin, A. McLeod, P. Rigaux, F. Jacquemard, and M. Sakai, "ASAP: a dataset of aligned scores and performances for piano transcription," in *International Society for Music Information Retrieval Conference (ISMIR)*, 2020, pp. 534–541.

[48] S. D. Peter, C. E. Cancino-Chacón, F. Foscarin, A. P. McLeod, F. Henkel, E. Karystinaios, and G. Widmer, "Automatic Note-Level Score-to-Performance Alignments in the ASAP Dataset," *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 2023.

[49] F. Foscarin, J. Schlüter, and G. Widmer, "Beat this! Accurate beat tracking without DBN postprocessing," Jul. 2024.

[50] M. Davies and S. Böck, "Evaluating the Evaluation Measures for Beat Tracking," in *Proceedings of the 15th International Society for Music Information Retrieval*. ISMIR, 2014.

[51] L. Ferreira-Paiva, E. Alfaro-Espinoza, V. M. Almeida, L. B. Felix, and R. V. A. Neves, "A Survey of Data Augmentation for Audio Classification," *Congresso Brasileiro de Automática - CBA*, vol. 3, no. 1, Oct. 2022.

[52] C. Drake, A. Penel, and E. Bigand, "Tapping in time with mechanically and expressively performed music," *Music Perception*, vol. 18, no. 1, pp. 1–23, 2000.

[53] S. Dixon, "A lightweight multi-agent musical beat tracking system," in *Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence*, ser. PRICAI'00. Berlin, Heidelberg: Springer-Verlag, 2000, pp. 778–788.

[54] J. S. Downie, "The Music Information Retrieval Evaluation eXchange (MIREX)," *D-Lib Magazine*, vol. 12, no. 12, Dec. 2006.

[55] M. McKinney, D. Moelants, M. Davies, and A. Klapuri, "Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms," *Journal of New Music Research*, vol. 36, pp. 1–16, Mar. 2007.

[56] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[57] A. T. Cemgil, "Bayesian music transcription," PhD Thesis, Radboud University Nijmegen, 2004.

[58] M. Goto and Y. Muraoka, "Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions," *Speech Communication*, vol. 27, no. 3-4, pp. 311–335, Apr. 1999.

[59] E. Large and J. Kolen, "Resonance And The Perception Of Musical Meter," *Connect. Sci*, vol. 6, Jan. 1998.