# Numerical sheet music analysis, L3 intership (CNAM / INRIA) 27/05 - 02/08

Sylvain Meunier
sylvain.meunier@ens-rennes.fr

Florent Jacquemard
florent.jacquemard@inria.fr

## I. Introduction

We present here some results regarding the analysis of tempo curve of musical performances, with score-based and scoreless approaches extending previously existing models.

The Music Information Retrieval (MIR) community focus on three ways to compute musical information. The first one is raw audio, either recorded or generated, encoded in .wav or .mp3 files. The computation is based on a physical understanding of signals, using audio frames and spectrum, and represents the most common and accessible type of data. The second is a more musically-informed format, that indicates mainly two parameters : pitch (ie the note that the listener hear) and duration, encoded within a .mid (or MIDI) file. Such a file can be displayed as a piano roll, that is a graph whose x-axis is time and y-axis is pitch (hence, the y-axis is discrete). The last way to encode musical information is the computed counterpart of sheet music. A sheet music is a way to write down a musical score, that is usually computed as a .music_xml file, mainly for display purposes. It comes with a symbolic and abstract notation for time, that only describes the length of events in relation to a specific abstract unit, called a beat, and the pitch of each event. This kind of data is actually the least common and accessible.

To actually play a sheet music, one needs a given tempo, usually indicated as the amoung of beat per minute (BPM). Therefore, the notion of tempo allows to translate symbolic notation (expressed in musical unit, eg : beats) to real time events (expressed in real time unit, eg : seconds). We will discuss later on a formal definition of tempo. However, tempo itself is insufficient to describe an actual performance of a sheet music, ie the sequence of real time events. Indeed, S. D. Peter, C. E. Cancino-Chacón, E. Karystinaios, and G. Widmer [1] present four parameters, among which tempo and articulation appear the most salient in contrast with velocity and timing. The latter represents the delay between the theorical real time onset according to the current tempo, and the actual onset heared in the performance. Even though such a delay is inevitable for neurological and biological reasons, those timings are usually overemphasized and understood as part of the musical expressivity of the performance.

In this study, we shall focus mainly on tempo estimation for a given performance recorded as a MIDI file, on both a local and global level.

## II. State of Art

Even though the community studies the four parameters, the hierarchy [1] exposed embodies quite well the importance within the litterature. O. F. B. Katerina Kosta Rafael Ramírez and E. Chew [2] present results pointing that, although velocities don't help to meaningfully estimate tempo, the latter allows to marginally upgrade velocity-related predictions. Actually, velocity appears to be more of a score parameter rather than a performance one : automatic learning methods trained on performances of a single piece showed much better results when asked to predict velocities employed by another performer on the same piece than when trained on other performances of the same performer.

Tempo and related works actually hold a prominent place in litterature. Direct tempo estimation was first computed based on probabilistic models (C. Raphael [3], E. Nakamura, N. Ono, S. Sagayama, and K. Watanabe [4], [5]), and physical / neurological models (E. W. Large and M. R. Jones [6], H.-H. Schulze, A. Cordes, and D. Vorberg [7]) ; before the community tried neural network models [2] and hybrids approaches (K. Shibata, E. Nakamura, and K. Yoshii [8]). As the majority of previous examples, we shall focus here on mathematically and/or musically explainable methods.

Since tempo needs a symbolic representation to be meaningful, one can consider transcription as a tempo-related work. We will keep this discussion for section V and VI.

However, note-alignement, that is matching each note of a performance with those indicated by a given score is a very useful preprocessing technique, especially for direct tempo estimation and further analysis, such as [9]–[11]. Two main methods are to be found in litterature : a dynamic programming algorithm, equivalent to finding a shortest path (M. Müller [12]), that can works on raw audio (.wav files) ; and a Hidden Markov Model (E. Nakamura, K. Yoshii, and H. Katayose [13]) that needs more formatted data, such as MIDI files.

In this report, we will present a few contributions :

- a justified proposition for a formal definition of tempo based on C. Raphael [3], [9] and P. Hu and G. Widmer [11] ; and some direct consequences
- a modification of E. W. Large and M. R. Jones [6] and H.-H. Schulze, A. Cordes, and D. Vorberg [7]
- an extension of G. Romero-García, C. Guichaoua, and E. Chew [14], to fit tempo estimation

## III. Score-based approaches

Since we chose to focus on MIDI files, we will represent a performance as a strictly increasing sequence of events $(t_n)_{n\in\mathbb{N}}$, each element of whose indicates the onset of the corresponding performance event. Such a definition is very close to an actual MIDI representation.

For practical considerations, we will stack together all events whose distance in time is smaller than $\varepsilon = 20$ ms. This order of magnitude, calculated by E. Nakamura, T. Nakamura, Y. Saito, N. Ono, and S. Sagayama [5] represents the limits of human ability to tell two rythmic events appart, and is widely used within the field [8]–[11], [13]–[16].

Likewise, a sheet music will be represented as a strictly increasing sequence of events $(b_n)_{n\in\mathbb{N}}$. In both of those definition, the terms of the sequence do not indicate the nature of the event (chord, single note, rest...). Moreover, in terms of units, $(t_n)$ corresponds to real onset, thus expressed in seconds, whereas $(b_n)$ corresponds to theorical or symbolic onsets, expressed in beats.

With those definitions, let us then formally define tempo $T(t)$ so that, for all $n \in \mathbb{N}$, $\int_{t_0}^{t_n} T(t)\,\mathrm{d}t = b_n - b_0$.
Appendix A shows that this definition is equivalent to : $\forall n \in \mathbb{N}, \int_{t_n}^{t_{n+1}} T(t)\,\mathrm{d}t = b_{n+1} - b_n$. However, tempo is only tangible (or observable) between two events *a priori*. We will then define the canonical tempo $T^*(t)$ so that :
$\forall x \in \mathbb{R}^+, \forall n \in \mathbb{N}, x \in [t_n, t_{n+1}[ \Rightarrow T^*(x) = \frac{b_{n+1}-b_n}{t_{n+1}-t_n}$.
The reader can verify that this function is a formal tempo according to the previous definition. From now on, we will consider the convention : $t_0 = 0$ (s) et $b_0 = 0$ (beat).

Even though there is a general consensus in the field as for the interest and informal definition of tempo, several formal definitions coexist within litterature : K. Shibata, E. Nakamura, and K. Yoshii [8] and E. Nakamura, N. Ono, S. Sagayama, and K. Watanabe [4] take $\frac{1}{T^*}$ as definition ; C. Raphael [3], [9] et P. Hu and G. Widmer [11] choose similar definitions than the one given here (approximated at the scale of a measure or a section for instance).

$T^*$ has the advantage to coincide with the tempo actually indicated on traditional sheet music (and therefore on .music_xml format), hence allowing a simpler and more direct interpretation of results.

### A. version naïve

- données (n)ASAP d'alignement
- tempo brute
- médiane fenêtre glissante

### B. modèles physiques

- Large: oscillateurs amortis
- TimeKeeper
- boom : LargeKeeper

## IV. Scoreless approaches

### A. principe: SoA quantification rythmique MIDI

2 papier

### B. approche estimateur

### C. approche quantifiée "spectrale" à la Gonzalo

- LR
- bidi (2 passes: LR + RL)
- RT : avec valeur initiale de tempo

### D. résultats évaluation (comparaison avec 3)

## V. Applications

- previous : metronaut, antescofo
- génération de données "performance" : pour data augmentation ou test robustesse (fuzz testing) aplanissement de tempo démo MIDI?
- transcription MIDI par parsing : pre-processing d'évaluation tempo (approche partie 4)
- analyse "musicologique" quantitative de performances humaines de réf. (à la Mazurka BL) données quantitives de tempo et time-shifts
- accompagnement automatique RT avec approche 4 RT ?

## VI. Conclusion & perspectives

- intégration pour couplage avec transcription par parsing (+ plus court chemin multi-critère)
- lien approche partie 4 "spectrale" avec Large (amortisseur) modification modèle Large : résultat théorique de convergence

## I. Appendix A

### A. Equivalence of tempo formal definitions

Let $n \in \mathbb{N}$. $\int_{t_0}^{t_n} T(t)\,\mathrm{d}t = \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} T(t)\,\mathrm{d}t$

Furthermore, $\int_{t_n}^{t_{n+1}} T(t)\,\mathrm{d}t = \int_{t_0}^{t_{n+1}} T(t)\,\mathrm{d}t + \int_{t_n}^{t_0} T(t)\,\mathrm{d}t =$

$\int_{t_0}^{t_{n+1}} T(t)\,\mathrm{d}t - \int_{t_0}^{t_n} T(t)\,\mathrm{d}t$.
We thus obtain the two implications, hence the equivalence.

### B. *BeatKeeper*

On cherche ici à déterminer une équation pour la période, en fusionnant les modèles E. W. Large and M. R. Jones [6] et J. D. Loehr, E. W. Large, and C. Palmer [17]. On reprend donc l'équation de la phase donnée par E. W. Large and M. R. Jones [6] : EQ1

On cherche à calculer : $T_n = \frac{1}{p_n} = \frac{\Phi_{n+1} - \Phi_n}{t_{n+1} - t_n}$ On considérant $\Phi_n$ comme le déphasage entre l'oscillateur de période $p_n$ et un oscillateur extérieur…

On a : 
$$\begin{aligned} \Phi_{n+1} - \Phi_n &= \frac{\Delta t_n}{p_n} - \eta_\Phi F(\Phi_n) \\ &= T_n \Delta t_n - \eta_\Phi F(\Phi_n) \\ &= T_n \frac{b_{n+1} - b_n}{T_n^*} - \eta_\Phi F(\Phi_n) \\ &= \Delta b_n \frac{T_n}{T_n^*} - \eta_\Phi F(\Phi_n) \end{aligned}$$

## II. Annexe B

## III. Annexe C

Posons tout d'abord quelques fonctions utiles.
On définit : $g : x \mapsto \min(x - \lfloor x \rfloor, 1 + \lfloor x \rfloor - x)$
On peut vérifier que $g : x \mapsto \begin{cases} x - \lfloor x \rfloor \text{ si } x - \lfloor x \rfloor \leq \frac{1}{2} \\ 1 - (x - \lfloor x \rfloor) \text{ sinon} \end{cases}$ et que $g$ est 1-périodique continue sur $\mathbb{R}$.

Ainsi, on a : $\varepsilon_T(a) = \max_{t \in T} g\left(\frac{t}{a}\right)$, donc en particulier, $\varepsilon_T$ est continue sur $R_+^*$.

On remarque de plus, pour $n \in \mathbb{N}^*, T \subset \left(\mathbb{R}_+^*\right)^n, a \in R_+^*$ : $\varepsilon_T(a) = a \varepsilon_{T/a}(1)$

### A. *Caractérisation des maximums locaux*

### B. *Caractérisation des minimums locaux*

Par continuité de $\varepsilon_T$, on est assuré de l'existence d'exactement un unique minimum local entre deux maximums locaux, qui est alors global sur cet intervalle.
Par la condition nécessaire précédente, il suffit donc, pour déterminer ce minimum local, de déterminer le plus petit élément parmi les points obtenus, contenus dans l'intervalle. On en déduit ainsi un algorithme en $\mathcal{O}\left(\#T^2 \frac{t^*}{\tau} \log\left(\#T \frac{t^*}{\tau}\right)\right)$ permettant de déterminer tous les minimums locaux accordés par le seuil $\tau$ fixé, sur l'intervalle $]2\tau, t_* + \tau[$

## IV. Glossary

### A. *Acronyms*

*MIR* – Music Information Retrieval: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do. 1

### B. *Definitions*

*articulation*. 1

*beat*:
Unité de temps d'une partition, le beat est défini par une signature temps, ou division temporelle, rappelée au début de chaque système. Bien que sa valeur ne soit *a priori* pas fixe d'une partition à une autre, ni même sur une même partition, la notion de beat est en général l'unité la plus pratique quant à la description d'un passage rythmique, lorsque la signature temps est adéquatement définie. 1, 3

*cadence*.

*chord*. 2

*measure*: Une mesure est une unité de temps musicale, contenant un certain nombre (entier) de beat. Ce nombre est indiqué par la time signature 2

*monophonic*.

*phrase*.

*polyphonic*.

*rest*. 2

*section*. 2

*tatum*: Résolution minimal d'une unité musicale, exprimé en beat. Bien que de nombreuses valeurs soit possible, la définition formelle d'un tatum serait la suivante : $\sup\{r \mid \forall n \in \mathbb{N}, \exists k \in \mathbb{N} : b_n = kr, r \in \mathbb{R}_+^*\}$. Pour des raisons pratiques, il arrive que le tatum soit un élément plus petit que la définition donnée, en particulier si cet élément est plus facilement expressible dans une partition, ou a plus de sens d'un point de vue musical. On notera dans la définition de l'ensemble donnée, k n'a pas d'unité, ce qui montre clairement que le tatum s'exprime en beat comme dit précédemment.

*tempo*:
Défini formellement p. 1 selon la formule : $T_n^* = \frac{b_{n+1} - b_n}{t_{n+1} - t_n}$. Informellement, le tempo est une mesure la vitesse instantanée d'une performance, souvent indiqué sur la partition. On peut le voir comme le rapport entre la vitesse symbolique supposée par la partition, et la vitesse réelle d'une performance. Le tempo est usuellement indiqué en beat par minute, ou bpm 1

*time signature*. 3

*velocity*. 1

## References

[1] S. D. Peter, C. E. Cancino-Chacón, E. Karystinaios, and G. Widmer, "Sounding Out Reconstruction Error-Based Evaluation of Generative Models of Expressive Performance," in *Proceedings of the 10th International Conference on Digital Libraries for Musicology*, 2023, pp. 58–66.

[2] O. F. B. Katerina Kosta Rafael Ramírez and E. Chew, "Mapping between dynamic markings and performed loudness: a machine learning approach," *Journal of Mathematics and Music*, vol. 10, no. 2, pp. 149–172, 2016, doi: 10.1080/17459737.2016.1193237.

[3] C. Raphael, "A Probabilistic Expert System for Automatic Musical Accompaniment," *Journal of Computational and Graphical Statistics*, vol. 10, no. 3, pp. 487–512, Sep. 2001, doi: 10.1198/106186001317115081.

[4] E. Nakamura, N. Ono, S. Sagayama, and K. Watanabe, "A Stochastic Temporal Model of Polyphonic MIDI Performance with Ornaments," *Journal of New Music Research*, vol. 44, no. 4, pp. 287–304, Oct. 2015, doi: 10.1080/09298215.2015.1078819.

[5] E. Nakamura, T. Nakamura, Y. Saito, N. Ono, and S. Sagayama, "Outer-Product Hidden Markov Model and Polyphonic MIDI Score Following," *Journal of New Music Research*, vol. 43, no. 2, pp. 183–201, Apr. 2014, doi: 10.1080/09298215.2014.884145.

[6] E. W. Large and M. R. Jones, "The dynamics of attending: How people track time-varying events," *Psychological Review*, vol. 106, no. 1, pp. 119–159, 1999, doi: 10.1037/0033-295X.106.1.119.

[7] H.-H. Schulze, A. Cordes, and D. Vorberg, "Keeping Synchrony While Tempo Changes: Accelerando and Ritardando," *Music Perception: An Interdisciplinary Journal*, vol. 22, no. 3, pp. 461–477, 2005, doi: 10.1525/mp.2005.22.3.461.

[8] K. Shibata, E. Nakamura, and K. Yoshii, "Non-local musical statistics as guides for audio-to-score piano transcription," *Information Sciences*, vol. 566, pp. 262–280, Aug. 2021, doi: 10.1016/j.ins.2021.03.014.

[9] "MazurkaBL: Score-aligned Loudness, Beat, and Expressive Markings Data for 2000 Chopin Mazurka Recordings." Accessed: Jun. 18, 2024. [Online]. Available: https://zenodo.org/records/1290763

[10] J. Hentschel, M. Neuwirth, and M. Rohrmeier, "The Annotated Mozart Sonatas: Score, Harmony, and Cadence," vol. 4, no. 1, pp. 67–80, May 2021, doi: 10.5334/tismir.63.

[11] P. Hu and G. Widmer, "The Batik-plays-Mozart Corpus: Linking Performance to Score to Musicological Annotations." Accessed: Jun. 18, 2024. [Online]. Available: http://arxiv.org/abs/2309.02399

[12] M. Müller, "MEMORY-RESTRICTED MULTISCALE DYNAMIC TIME WARPING," Accessed: Jun. 18, 2024. [Online]. Available: https://www.academia.edu/25724042/MEMORY_RESTRICTED_MULTISCALE_DYNAMIC_TIME_WARPING

[13] E. Nakamura, K. Yoshii, and H. Katayose, "Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment," 2017. Accessed: Jun. 18, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Performance-Error-Detection-and-Post-Processing-for-Nakamura-Yoshii/37e9f5e23cada918c2b8982d71a18972140d9d5a

[14] G. Romero-García, C. Guichaoua, and E. Chew, "A Model of Rhythm Transcription as Path Selection through Approximate Common Divisor Graphs," May 2022. Accessed: Jun. 19, 2024. [Online]. Available: https://hal.science/hal-03714207

[15] F. Foscarin, A. Mcleod, P. Rigaux, F. Jacquemard, and M. Sakai, "ASAP: a dataset of aligned scores and performances for piano transcription," Oct. 2020. Accessed: Jun. 18, 2024. [Online]. Available: https://cnam.hal.science/hal-02929324

[16] S. D. Peter *et al.*, "Automatic Note-Level Score-to-Performance Alignments in the ASAP Dataset," vol. 6, no. 1, pp. 27–42, Jun. 2023, doi: 10.5334/tismir.149.

[17] J. D. Loehr, E. W. Large, and C. Palmer, "Temporal coordination and adaptation to rate change in music performance," *Journal of Experimental Psychology. Human Perception and Performance*, vol. 37, no. 4, pp. 1292–1309, Aug. 2011, doi: 10.1037/a0023102.