Analyse de partitions musicales numériques
Rapport de stage L3
27/05 - 02/08

Sylvain Meunier
sylvain.meunier@ens-rennes.fr

I. Introduction

Nous présentons ici quelques résultats obtenus autour de la notion de tempo en informatique musicale, ou Music Information Retrieval (MIR), notamment dans le domaine d'estimation du tempo. En ce qui concerne le temps réel, l'une des problèmatiques les plus saillante est celle de l'accompagnement automatique d'un soliste [1], [2]. L'objectif d'un tel modèle est alors de synchroniser la lecture d'une partition par une machine avec le jeu d'au moins un humain. Récemment, une approche reprenant [2] a été développé pour un usage commercial.[1] Un pré-calcul non trivial nécessaire à une bonne approche consiste à "aligner" les notes d'une performance (jouées par un humain et perçues par le système) avec celle d'une représentation symbolique interne (partition, ou fichier midi). Ce problème d'alignement de séquences est très similaire à son équivalent textuel (on peut de fait voir une partition ou sa réalisation comme un mot sur un alphabet musical). On dispose de deux outils pour résoudre ce problème : un algorithme de programmation dynamique revenant à chercher un plus court chemin [3] (qui fonctionne également sur des fichiers sons "bruts", .wav notamment) et un modèle de Markov caché [4] (qui nécessite des données plus formattées, .mid en particulier).

Caser quelque part les arbres de rythme et l'autocorrélation (quel nakamura ?) pour la transcription (rythmique)

Parmi les quatre paramètres qui caractérisent une performance, [5] en distingue deux : le tempo, et l'articulation ; mettant de côté le *timing* et les nuances. Cette hiérarchie est assez représentative des préoccupations de la communauté : [6] présente des résultats indiquant que si les nuances ne permettent pas d'estimer de façon pertinente le tempo, connaître celui-ci permet d'améliorer les prédictions relatives aux nuances, qui sont caractéristiques d'une composition plus que d'un style de jeu : les méthodes d'apprentissage automatiques employées entraînées sur des performances d'une même pièce parvenaient bien mieux à prédire les nuances utilisées par un pianiste inconnu sur cette pièce, que celles entraînées sur d'autres performances d'autres pièces de ce même pianiste. Retour sur le tempo... revoir éventuellement cette hiérarchie [et nous chercherons à lier articulation et tempo]. Ce dernier possède une place prépondérante dans la littérature. Si des modèles probabilistes [1], [7], [8] sont d'abord développés, ainsi que des modèles physiques, voire neurologiques [9], [10] ; la communauté s'est tournée plus récemment vers des approches comprenant des réseaux de neuronnes [6], ce qui donne lieu à l'apparition d'approches hybrides [11]. A l'instar de la plupart des exemples cités, on se concentrera ici sur des méthodes explicables mathématiquement et/ou musicalement.

II. Estimation du tempo

A. *Présentation du formalisme utilisé*

On dispose de deux façons principales de représenter une performance informatiquement : un fichier audio brute en format .wav par exemple, ou bien un fichier midi plus symbolique. Afin de simplifier les algorithmes, nous ne considèrerons ici que des entrées sous forme de fichiers midi. On modélise alors une performance comme une suite strictement croissante d'évènements $(t_n)_{n\in\mathbb{N}}$, dont chaque élément indique la date de l'évènement associé. Cette modélisation coïncide presque avec le contenu d'un fichier midi. Pour des considérations pratiques, on regroupera ensemble les évènements distants dans le temps de $\varepsilon = 20$ ms, ordre de grandeur calculé par [8] dont la valeur correspond à la limite de la capacité humaine à distinguer deux évènements rythmiques. Cet ordre de grandeur est largement utilisé dans le domaine [4], [12]–[17], [11]. De façon similaire, on modélise une partition comme une suite strictement croissante d'évènements $(b_n)_{n\in\mathbb{N}}$. On notera que dans ces deux définitions, les éléments de la suite indique certes un évènement, mais pas sa nature. Il peut donc notamment s'agir d'un accord, d'une note seule, ou d'un silence. En termes d'unité, on notera que (t_n) désigne des dates réelles, en secondes par exemple ; alors que (b_n) désigne des dates théoriques, exprimées en beat, unité de temps musicale. On peut alors définir formellement le tempo $T(t)$ de sorte que, pour tout $n \in \mathbb{N}$, $\int_{t_0}^{t_n} T(t)\,\mathrm{d}t = b_n - b_0$.

[1] https://metronautapp.com/

On montre en Annexe A que cette définition est équivalente à : $\forall n \in \mathbb{N}, \int_{t_n}^{t_{n+1}} T(t)\,\mathrm{d}t = b_{n+1} - b_n$.
Or, le tempo n'est tangible (ou observable) qu'entre deux évènements *a priori*. On définira donc le tempo canonique $T^*(t)$ de sorte que :
$\forall x \in \mathbb{R}^+, \forall n \in \mathbb{N}, x \in [t_n, t_{n+1}[\Rightarrow T^*(x) = \frac{b_{n+1}-b_n}{t_{n+1}-t_n}$.
On peut alors s'assurer que cette fonction respecte bien la condition énoncée précédemment. Par convention, on prendra dans la suite de ce rapport : $t_0 = 0$ et $b_0 = 0$.

Si le domaine présente un consensus global quant à l'intérêt et la définition informelle de tempo, de multiples définitions formelles coexistent dans la littérature : [11] et [7] prennent pour définition $\frac{1}{T^*}$; [1], [15] et [14] choisissent des définitions proches de celle donnée ici (plus ou moins approximée à l'échelle d'une mesure ou d'une section par exemple). T^* a l'avantage de coïncider avec le tempo indiqué sur une partition, et donc de permettre une interprétation plus directe des résultats. Par ailleurs, nous avons déjà présenté une justification de la formule permettant de calculer le tempo canonique.

B. Approche naïve

Etant donné ce formalisme, on peut alors construire un algorithme glouton calculant le tempo entre les instants n et $n+1$ d'une performance lorsque la partition est connue, selon la formule : $T_n^* = \frac{b_{n+1}-b_n}{t_{n+1}-t_n}$, dont le lecteur peut s'assurer de l'homogénéité.
On donne en figure 1 quelques résultats donnés par cette approche dans différentes situations :

- situation théorique parfaite
- situation réelle approchant une situation théorique, par l'ajout de perturbations
- situation réelle (sur Mozart, nom_de_loeuvre)

C. Approches existantes

1) *Large et Jones:*
L'approche de Large et Jones [9] considère un modèle neurologique simplifié, dans lequel l'écoute est fondamentalement active, et implique une synchronisation entre des évènements extérieurs (la performance) et un oscillateur interne, plus ou moins complexe selon la forme supposée de ces premiers. Le modèle consiste en deux équations pour les paramètres internes :

$$\Phi_{n+1} = \left[\Phi_n + \frac{t_{n+1}-t_n}{p_n} - \eta_\Phi F(\Phi_n)\right] \bmod 1 \qquad (1)$$

$$p_{n+1} = p_n\big(1 + \eta_p F(\Phi_n)\big) \qquad (2)$$

Ici, Φ_n correspond à la phase, ou plutôt ou déphasage entre l'oscillateur et les évènements extérieurs, et p_n désigne sa période.
Ce modèle initial est ensuite revu pour y incorporer une notion d'attention *via* le paramètre κ, non constant au cours du temps. Les formules restent alors les mêmes, en remplaçant F par $F : \Phi, \kappa \to \frac{\exp(\kappa\cos(2\pi\Phi))}{\exp(\kappa)}\frac{\sin(2\pi\Phi)}{2\pi}$
Si ce modèle se comporte très bien en pratique, a été validé par l'expérience dans [9], et reste encore utilisé dans la version présentée ici [18], l'étude théorique du comportement du système n'en est pas aisée [10], même dans des cas simples, notamment en raison de l'expression de la fonction F.

2) *TimeKeeper:*
Dans un soucis de simplification du modèle, [10] présente TimeKeeper, qui peut être perçu comme une simplification de l'approche précédente, valide dans le cadre théorique d'un métronome présentant de faibles variations de tempo. On peut toutefois voir une presque équivalence entre les deux modèles [19]. On montre en figure XXX une comparaison dans différents contextes des trois approches citées jusqu'à présent, où on note une stabilité saillante du modèle d'oscillateur.

D. Contributions

1) *BeatKeeper:*
Un premier objectif a été de fusionner les approches de [9] et [19] afin d'essayer d'obtenir des garanties théoriques sur le modèle résultant. On montre en Annexe A que l'on obtient alors le système composé des deux équations suivantes :

$$\Phi_{n+1} = \left[\Phi_n + \frac{t_{n+1}-t_n}{p_n} - \eta_\Phi F(\Phi_n)\right] \bmod 1 \qquad (3)$$

$$p_{n+1} = p_n \frac{1}{1 - \frac{p_n \eta_\Phi F(\Phi_n, \kappa_n)}{t_{n+1}-t_n}} \qquad (4)$$

On notera que ce modèle a l'avantage de contenir un paramètre de moins que celui de Large. On remarque également que, pour $\Delta_t = t_{n+1} - t_n \gg p_n\eta_\Phi F(\Phi_n, \kappa_n)$ dans (4), on obtient : $p_{n+1} = p_n\left(1 + \frac{p_n}{\Delta_t}\eta_\Phi F(\Phi_n, \kappa_n)\right)$. Quitte à poser $\eta_p = \frac{p_n}{\Delta_t}\eta_\Phi$ on retrouve (2).
Les modèles sont donc équivalents sous ces conditions. En pratique (voir Annexe B), on obtient des résultats très similaires à [9], avec un paramètre constant en moins. Bien que ce modèle n'offre guère plus de garanties *a priori* que [9], il est toutefois nettement plus aisé d'en faire l'analyse par rapport à un tempo canonique. On peut ainsi montrer que :

$$\begin{aligned}\alpha_{n+1} &= \left(\alpha_n - \eta_\Phi \frac{F(\Phi_n)}{\Delta b_n}\right)\frac{T_n^*}{T_{n+1}^*} \\ &= \alpha_n \frac{T_n^*}{T_{n+1}^*} - \eta_\Phi \frac{F(\Phi_n)}{\Delta t_n T_{n+1}^*}\end{aligned} \qquad (5)$$

où $\forall n \in \mathbb{N}, \alpha_n = \frac{T_n}{T_n^*}$.
Par ailleurs, on montre en Annexe B que ce modèle possède les mêmes garanties théoriques que [9] dans une situation idéalisée simple.

2) *TempoTracker:*
Après l'approche précédente,

3) *Analyse des résultats:*

On définit le spectre d'un algorithme : blabla

On utilise la mesure d'un spectre $S \neq \emptyset$, qui correspond en quelque sorte à l'écart-type, de sorte que

$$m(S, \Delta) = \max_{d \in \mathcal{C}} \frac{\#\{s \in S, |s, d| \leq \Delta\}}{\#S} \tag{6}$$

Alors, Δ correspond à la moitié de la largeur autorisée d'un pic, donc la "précision" de la mesure. On a dans notre cas $\Delta \leq 1$. On peut vérifier que cette mesure est invariante par rotation du spectre (multiplication de toutes les valeurs du spectre par une constante, puis normalisation du résultat), ainsi que de choix de l'intervalle de normalisation ; et que $0 < m(S, \Delta) \leq 1$, avec égalité si et seulement si S est constitué d'une raie unique avec la précision Δ. Avec la convention $m(\emptyset, \Delta) = 0$, on peut même affirmer que $m_\Delta : S \to m(\Delta, S)$ est une norme sur les spectres.

En effet, on a : $m_\Delta(S) = 0 \Leftrightarrow S = \emptyset$ par la convention précédente, et reste à vérifier que $m_\Delta(S \cup S') \leq m_\Delta(S) + m_\Delta(S')$

III. Applications

I. Annexe A

A. Equivalence des définitions du tempo

Soit $n \in \mathbb{N}$, on a :

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

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

On obtient ainsi les deux implications.

B. BeatKeeper

On cherche ici à déterminer une équation pour la période, en fusionnant les modèles [9] et [19]. On reprend donc l'équation de la phase donnée par [9] :

$$\Phi_{n+1} = \left[\Phi_n + \frac{t_{n+1} - t_n}{p_n} - \eta_\Phi F(\Phi_n)\right] \bmod 1 \tag{7}$$

On cherche à calculer : $T_n = \frac{1}{p_n} = \frac{\Phi_{n+1} - \Phi_n}{t_{n+1} - t_n}$ On considérant Φ_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 : $\varepsilon_T(a) = \max_{t \in T} g(\frac{t}{a})$, donc en particulier, ε_T est continue sur R_+^*.

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

A. Caractérisation des maximums locaux

B. Caractérisation des minimums locaux

Par continuité de ε_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 τ fixé, sur l'intervalle $]2\tau, t_* + \tau[$

IV. Glossaire

potato: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

A. Acronymes

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

B. Définitions

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, 4

cadence.

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

monophonique.

nuances. 1

phrase.

polyphonique.

section. 2

tatum: Résolution minimal d'une unité musicale, exprimé en beat. Bien que de nombreuses valeurs soit possi-

ble, 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, 2

signature de temps. 3

References

[1] 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.

[2] A. Cont, F. Jacquemard, and P.-O. Gaumin, "Antescofo à l'avant-garde de l'informatique musicale," *Interstices*, Nov. 2012, [Online]. Available: https://inria.hal.science/hal-00753014

[3] 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

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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

[13] 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.

[14] 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

[15] "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

[16] 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.

[17] 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

[18] E. W. Large *et al.*, "Dynamic models for musical rhythm perception and coordination," *Frontiers in Computational Neuroscience*, vol. 17, May 2023, doi: 10.3389/fncom.2023.1151895.

[19] 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.