UN JEU DE DONNÉES POUR L'ENTRAÎNEMENT DE SYSTÈMES DE SYNCHRONISATION MUSICIEN/MACHINE

Serge Lemouton
STMS – IRCAM, CNRS
serge.lemouton@ircam.fr

Jean-Louis Giavitto
STMS – CNRS, IRCAM
giavitto@ircam.fr

Miller Puckette
STMS – UCSD, IRCAM
msp@ucsd.edu

RÉSUMÉ

Le suivi de partition, défini comme l'alignement en temps réel d'un flux audio sur une référence temporelle symbolique (partition, cue-list) reste difficile à rendre fiable en situation réelle dans le contexte de la musique contemporaine. Les modes de jeu étendus, les timbres bruités et les écarts interprétatifs mettent en défaut des approches historiques, fondées sur des modèles probabilistes même si des applications destinées à la pratique tonale montrent qu'un usage stable est possible dans des cadres stylistiquement contraints.

La situation évolue avec le développement de nouvelles techniques (comme l'apprentissage profond) qui renouvelle les représentations audio et améliore la robustesse des systèmes. Par ailleurs, le problème se généralise : il s'agit plus généralement d'aligner en temps-réel un ou plusieurs médias à un scénario temporel prédéfini ouvrant des usages scéniques élargis.

Cependant, l'évaluation demeure un verrou. Les jeux de données publics offrant une "vérité terrain" (Ground Truth, GT) sont rares, coûteux à produire et hétérogènes, ce qui limite les comparaisons reproductibles. Nous proposons donc une approche pragmatique : constituer un corpus d'exécutions réelles assorties de traces d'alignement issues de conditions de concert (suivi hybride automatique–manuel). Ces traces ne constituent pas une GT stricte, mais une référence opérationnelle suffisante pour le benchmarking inter-systèmes.

1. INTRODUCTION

Le *suivi de partition* (ou *score following*) désigne l'alignement en temps réel d'un flux observé — typiquement audio, mais aussi MIDI ou signaux issus de capteurs — sur une représentation symbolique temporelle telle qu'une partition musicale notée et encodée informatiquement. Introduit dès les années 1980 dans les articles fondateurs de Barry Vercoe et Roger Dannenberg [18, 6] ce problème a donné lieu à différents modèles, tout d'abord de type pattern matching puis probabilistes, notamment formalisés à base de modèle de Markov caché (HMM) et de programmation dynamique [16].

Dans le répertoire tonal classique, ces approches ont atteint un degré de maturité suffisant pour permettre des usages pédagogiques et grand public. Cependant, dans le contexte de la musique contemporaine — textures bruitées, micro-intervalles, structures temporelles non métriques, modes de jeu étendus — le suivi de partition reste difficilement utilisable "en routine". Les hypothèses implicites de nombreux modèles (stabilité spectrale, correspondance note-à-note, métrique régulière) y sont fréquemment violées. Cette limite n'est pas seulement algorithmique ; elle est aussi liée aux conditions réelles d'exécution scénique, où l'incertitude, l'acoustique et l'interprétation introduisent une variabilité forte. Aujourd'hui, des œuvres emblématiques du suivi de partition (telles que *Anthèmes 2* de Pierre Boulez ou *Pluton* de Philippe Manoury) comportent toujours des passages qui nécessitent des déclenchements "à la main" de la part de l'interprète de la partie électronique (réalisateur.ice en informatique musicale).

Il semble bien que, comme l'exprimait déjà Barry Vercoe il y a plus de quarante-cinq ans, nous n'ayons toujours pas "compris suffisamment bien la dynamique d'un ensemble de musiciens humain pour être capable de remplacer arbitrairement un membre de l'ensemble par un musicien synthétique (c'est à dire un modèle informatique) de manière à ce que les autres membres présents sur scène ne puissent pas faire la différence" ("to understand the dynamics of live ensemble performance well enough to replace any member of the group by a synthetic performer (i.e. a computer model) so that the remaining live members cannot tell the difference" [18]). Si on a parfois l'impression que les choses évoluent très vite dans le domaine des innovations liées aux applications de la technologie à la musique, ce problème posé il y a maintenant plus de quarante années par Barry Vercoe et Roger Dannenberg n'a pas trouvé de solution convaincante.

Cet article n'est pas le lieu pour un recensement exhaustif et historique de toutes les expériences de suivi de partition menées depuis les travaux pionniers des années 80. On pourra trouver une liste assez complète dans [14] ou [12, pp. 35-41]. La situation évolue néanmoins sous l'effet de l'amélioration de techniques existantes [8] et de nouvelles approches fondées sur l'apprentissage qui proposent des représentations audio-symboliques apprises et une robustesse accrue aux variations de timbre et de contexte [15].

Par ailleurs, le problème se généralise : il ne s'agit plus uniquement d'aligner un signal audio à une partition, mais plus largement d'aligner un média (MIDI, audio, geste,

vidéo) à un scénario temporel (partition, cue-list, déroulé, synopsis, ou même une instance spécifique du media d'entrée lui-même servant alors de référence). Cette extension ouvre des usages nouveaux, par exemple le couplage d'un système de suivi à un dispositif de capture gestuelle afin de suivre un chef d'orchestre ou un instrumentiste, la partition devenant un déroulé événementiel.

Toutefois, un verrou méthodologique subsiste : l'évaluation. Les campagnes comparatives ont montré l'importance de jeux de données annotés pour mesurer les performances [13], mais les corpus publics offrant une "vérité terrain" précise et exploitable demeurent rares, en particulier pour la musique contemporaine. Produire ces alignements est coûteux, requiert une expertise musicale fine et reste sujet à variabilité inter-annotateurs. De plus, les jeux de données disponibles reflètent souvent des hypothèses stylistiques restrictives, limitant la portée des comparaisons inter-systèmes.

Une évaluation strictement quantitative du suivi de partition, fondée par exemple sur une valeur [1] calculée à partir des écarts temporels entre onsets détectés et onsets de référence, bien qu'absolument nécessaire, ne constitue cependant pas une mesure exhaustive de la performance d'un système. Ce score suppose implicitement que les événements sont définis de manière univoque et que leur localisation temporelle est stable. Or cette hypothèse est souvent mise en défaut dans les pièces de concert, par exemple dans les contextes vocaux ou dans des modes de jeu bruités. Dans le chant, l'attaque peut être progressive, dépendre de l'articulation propre à chaque interprète, ou être précédée d'éléments consonantiques dont la frontière avec la voyelle n'est pas objectivement tranchée. Dans des textures bruitées ou étendues, l'enveloppe d'énergie peut être instable, non stationnaire et difficilement segmentable en événements discrets. La notion même d'"onset de référence" devient alors partiellement conventionnelle. De plus, dans le contexte d'une performance en direct, le système ne dispose pas du signal futur ; l'estimation d'un onset repose donc sur une information partielle, ce qui introduit un décalage structurel par rapport à une annotation réalisée a posteriori.

Dans ces conditions, la performance ne peut être réduite à une métrique événementielle. Il convient de compléter l'évaluation quantitative par une comparaison qualitative des systèmes, intégrant leur comportement global : stabilité temporelle, gestion des ambiguïtés, capacité de récupération après décrochage, et, plus largement, adéquation musicale en situation réelle. Cette dimension qualitative permet d'apprécier la musicalité et la robustesse opérationnelle du système, au-delà d'une simple concordance locale d'onsets.

Dans cet article, nous défendons l'idée que l'amélioration des algorithmes doit aller de pair avec la constitution de références d'évaluation adaptées aux usages scéniques contemporains. Nous proposons de constituer un corpus d'exécutions réelles assorties de traces d'alignement issues de conditions de concert, incluant des phases de suivi hybride (automatique et interventions manuelles). Ces traces ne constituent pas une GT absolue, mais une référence opérationnelle suffisamment robuste pour permettre un benchmarking comparatif et qualitatif entre systèmes. L'objectif est ainsi double : clarifier le statut des données d'alignement (en quoi et comment peuvent-elle servir de référence dans l'évaluation d'un système de suivi) et poser les bases d'une évaluation plus réaliste des systèmes de suivi dans des contextes artistiques exigeants.

1 . Telle que, par exemple, la *F-mesure* ou *F-score* qui fusionne la *précision* et le *rappel* en une seule note pour évaluer l'efficacité globale d'un modèle de classification.

Piece name	Composer	Instrument	Files	Duration	Events
Explosante-Fixe	Boulez	Flute	47	17:10	2022
Violin Sonatas	Bach	Violin	3	13:50	3996
K. 370	Mozart	Clarinet	4	14:44	2710
Dorabella	Mozart	Voice	4	01:44	229
Total			58	47:38	8957

Table 1. MIREX06 Score Following Reference Database

2. JEUX DE DONNÉES EXISTANTS

En 2006, Arshia Cont a soumis un jeu de donnée à MIREX (Music Information Retrieval Evaluation eXchange Competition) [13] dans la catégorie "Real-time Audio to Score Alignement a.k.a Score Following". Cette compétition sera ouverte jusqu'en 2017, date à laquelle elle disparaît de MIREX et n'est plus accessible en ligne depuis longtemps. Les fichiers contenus dans la base de donnée destinée à évaluer les systèmes de suivi de partition en compétition sont listés dans la Table 1.

« Malheureusement, la procédure d'évaluation n'est pas clairement exprimée (les liens vers la procédure étant aujourd'hui inaccessibles), elle a trop évolué depuis les premières années pour que les premières évaluations menées entre 2006 et 2009 soient comparables à celles d'aujourd'hui. Très peu d'auteurs ont soumis leurs résultats. » [2]

Aujourd'hui, on peut trouver sur internet de nombreuses base de données sonores et musicales qui pourraient être utilisées pour entraîner, tester et/ou évaluer de nouvelles stratégies de synchronisation. par exemple :

- `https://dezrann.net/corpora` [1]
- `https://github.com/CPJKU/msmd` [7]
- `https://github.com/bytedance/GiantMIDI-Piano` [11]
- `https://github.com/flippy-fyp/QualScofo` [12]
- `https://magenta.tensorflow.org/datasets/maestro#v300` [9]

2 . "Unfortunately, the evaluation procedure is rather opaque (all links to sample data in the evaluation are now inaccessible), and it has evolved a lot in the early years to render early evaluation results (these were run on very few systems from 2006 to 2009 anyway) incomparable to those of later evaluations. Throughout the running of the evaluation, very few authors submitted their algorithms. From 2018 to 2020 no submissions were received." [12, pp. 41-42].

3. EVALUATION DE LA QUALITÉ DU SUIVI DE PARTITION

3.1. Métriques d'évaluation des systèmes de suivi symbolique

Dans l'article de 2007 [4], Arshia Cont et al. posent les bases d'une mesure de la qualité d'un système de synchronisation musicale permettant d'évaluer les différentes stratégies et algorithmes d'alignement en temps réel ("assesment metrics") :

- l'*erreur* correspond à la différence entre le temps détecté et le temps de référence de chaque occurrence d'événement
- la *latence*, calculée comme la différence entre l'instant où l'événement est reconnu et celui où il est reporté par le système, correspondant au temps que le système met à détecter l'attaque d'une note plus la latence du système audio
- l'*offset*, la somme de l'erreur et de la latence
- le nombre de notes non détectées (*missed*)
- le nombre de notes non alignées (c'est à dire au delà d'un certain écart temporel arbitraire fixé à 300ms)
- le ratio d'événements non détectés (*miss rate*)
- le taux de non alignement (*misaligned rate*) avec une erreur supérieure à un seuil
- le nombre d'événements avant que le suivi ne décroche définitivement (*piece completion*)
- la latence moyenne pour tous les événements correctement alignés (en dessous du seuil) (*average latency*)
- la moyenne des offsets pour tous les événements alignés, qui correspond à une mesure de la réactivité du système (*average absolute offset*)
- la moyenne des valeurs absolue des erreurs exprime l'imprécision du système (*average imprecision*)
- la déviation standard des erreurs pour tous les événements alignés, qui exprime l'étalement des erreurs (*variance of error*)

Ces mesures peuvent être remises en question dans le cas précis de la synchronisation homme/machine en contexte de concert, ces métriques ayant d'ailleurs été dès le début contestées. Une formulation plus générale des métriques d'évaluation de l'alignement entre un enregistrement et une référence discrète a été proposée dans [17]. Ces métriques sont destinées à un alignement offline mais pourraient être appliqués à un contexte de synchronisation en temps réel.

Ce que les métriques proposées dans l'article [4] ne mesurent pas c'est la robustesse du système d'alignement aux variations des observations. Ces variations inévitables sont pourtant inhérentes aux situations dans lesquelles on veut utiliser concrètement ces systèmes : celui du concert live, qui comporte de nombreux aléas. Un système qu'on peut utiliser dans de telles circonstances doit résister non seulement aux "fausses notes" mais aussi à des variations du signal d'entrée quel qu'il soit. On suggère donc d'introduire au cours de l'évaluation d'un système des variations dans le signal de référence (bruit ajouté, filtrage, dégradation, ...). On peut pour cela utiliser des techniques d'augmentation de donnée couramment utilisées pour l'entraînement dans le domaine des jeux d'apprentissage en apprentissage machine.

Les mesures recensées dans l'article de 2007 permettent d'évaluer et de comparer des systèmes de suivi de partition mais semblent être liés au paradigme d'un suivi de partition déterministe de type pattern matching, en d'autres termes, cette métrique est elle généralisable et appropriée à d'autres approches du problème ?

La multiplication de ces métriques nous semble être le signe de la difficulté à évaluer la performance d'un système de suivi. Idéalement, un bon système de suivi de partition, un interprète synthétique ("synthetic performer"), devrait exhiber les mêmes comportement qu'un « bon musicien » auquel nous prêterons les qualités suivantes :

- Il est capable de s'adapter aux erreurs des autres (et rattraper les siennes !)
- Il réagit instantanément (ou avec une latence inférieure au seuil de perception)
- Il est capable d'anticiper (de jouer même avant le temps)
- Il a un sens du rythme (Rythmic consistency, tempo, swing, ...)
- Il s'adapte musicalement aux aléas du concert live (fausse note et changement de tempo des autres musiciens, etc.)
- Son comportement est robuste : face aux variations de jeux, sa réponse est suffisamment cohérente et prévisible par les autres musiciens.

Ces propriétés sont nécessaires pour « jouer ensemble » et s'adapter aux aléas de la performance. On voit à travers ces différents points qu'il est difficile de réduire l'évaluation d'un système de suivi à la précision de la datation des événements détectées. Par exemple, la plupart des méthodes d'évaluation n'incluent pas la notion de mesure du tempo qui semble pourtant essentielle dans ce contexte. Aussi, même si des évaluations quantitatives peuvent permettre d'améliorer des algorithmes ou des systèmes complets d'accompagnement automatique, rien ne pourra remplacer une évaluation subjective par des musiciens, experts ou non, par des réalisateurs en informatique musicale (qui sont les principaux utilisateurs desdits systèmes) ou par des auditeurs. Dans ce cadre des protocoles expérimentaux restent largement à définir.

3.2. Comment évaluer les approches de suivi continu ?

La situation pointée ci-dessus s'aggrave dans le cas du suivi audio-audio ou du suivi de geste, où la notion d'événement n'est pas explicite.

On utilise aujourd'hui d'autres approches à cette question de l'alignement en temps réel entre une partition et un signal sonore, qui ne reposent plus sur une modélisation de la partition symbolique sous forme discrète (*notes*) mais sur un alignement entre le signal sonore et une description continue d'une version de référence (Fig. 1) de la partition pré-enregistrée (*geste*) dans lequel la mise en correspondance se fait en continu dans le temps, entre un

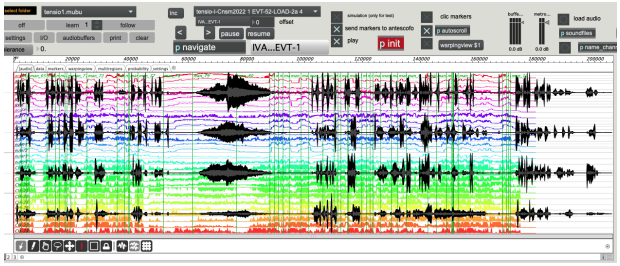

Figure 1. Exemple d'alignement continu pour le suivi du quatuor à corde *Tensio* de Philippe Manoury

flux audio en direct et un enregistrement de référence, sans faire appel à une partition écrite. Bien que cela soit assez différent de la méthode existante basée sur les notes, c'est une approche à base de HMM qui est utilisée, comme dans le suivi de partition combinatoire, mais les états de notre chaîne de Markov cachée sont des points temporels dans le modèle préenregistré (au lieu de notes) et des observations audio (correspondant à une fenêtre temporelle provenant du ou des instruments en direct, au lieu de descripteurs spectraux). Cette approche s'appuie sur celle adoptée depuis [14] par Bevilacqua et al. [2, 3]. Les métriques précédentes, dont on a déjà pointé le caractère partiel pour évaluer la réponse globale du système, sont dans ce contexte, largement inapplicables. On peut les généraliser, mais on peut adresser les mêmes critiques que précédemment à ces extensions.

4. UN CORPUS DE RÉFÉRENCE OPÉRATIONNEL

4.1. Evaluation à partir de données de "vérité-terrain"

Face à ces difficultés, nous proposons une approche pragmatique qui vise à compléter les évaluations quantitatives sur les données GT dont on dispose, par la comparaison qualitative des systèmes sur un corpus de pièces correspondant à des données réelles issues de concert.

En effet, il est à la fois important d'adresser des pièces longues (par exemple certains systèmes de suivi font structurellement plus d'erreurs au fur et à mesure qu'on avance dans la pièce) et de confronter les systèmes à plusieurs réalisations en vraie grandeur pour évaluer la pertinence musicale des comportement autre que la détection (rattrapage sur des fausses notes, robustesse des comportements face aux variations, etc.).

Les traces de références contenues dans le jeu de données proposé sont issues de concerts et donc validées par le concert. Elle peuvent servir de "vérité terrain" même si certains passages correspondent à des déclenchements manuels. Ces données de référence consistent en des fichiers sons (enregistrements en situation de concert ou de répétition) annotés par des marqueurs temporels correspondant aux points de déclenchement ou de synchronisation, et en des fichiers antescofo utilisés pour le suivi et la partition musicale (cf. Table 1).

fichiers sons	aiff ou wav
marqueurs	text (csv)
partition antescofo	text
partition musicale	midi, musicxml ou pdf

Table 1. formats de fichiers.

compositeur	titre	instrument	comment
S. Blondeau	*urphanomenon IIb*	piano	+ midi
P. Manoury	*Pluton*	piano	4 versions
Y. Maresz	*soli*	piano	
S. Blondeau	*Des mondes ...*	quatuor	
P. Manoury	*Partita 1*	alto	6 versions
P. Manoury	*Partita 2*	violon	
P. Manoury	*Tensio*	quatuor	2 versions
P. Boulez	*...explosante-fixe...*	flute	6 versions

Table 2. contenu du jeu de données.

4.2. Constitution du corpus

La base de données est accessible en ligne à l'adresse `https://forge-2.ircam.fr/POW/sxxi`. Elle con- tient des fichiers sons annotés par des marqueurs temporels et des partitions antescofo afin de tester et d'améliorer les stratégies de suivi de partition.

Le jeu de données créé par Arshia Cont pour les compétitions MIREX évoqué précédemment peut être intégrée à cette nouvelle base de données, qui pourrait être considérée comme une extension de celle-ci.

SXXI est un environnement de test pour des stratégies de synchronisation hybrides entre un signal audio musical et son accompagnement électroacoustique. Pour cela on combine des méthodes d'anticipation entre alignement discret (notes antescofo) et continu (geste xmm).

La base de données SXXI contient des enregistrements annotés de plusieurs versions d'exécutions réelles de pièces de musique du répertoire de l'IRCAM (Cf. Table 2). Par exemple, là où la base MIREX contenait une version de la flûte solo de *...explosante-fixe...* nous avons collecté cinq versions jouées par trois flûtistes différents. De même pour *Pluton*, le jeu de donnée contient les interprétations de trois pianistes différents.

Cette base de données est destinée à être enrichie au cours du temps, on encourage donc les utilisateurs de systèmes de suivi de partition (compositeurs, interprètes, réalisateurs en informatique musicale) à partager leurs enregistrements de répétitions annotés et accompagnés des partitions musicales. Dans le respect du droit d'auteur et des droits voisins, certaines parties de la base de données pourront être restreintes exclusivement à des usages de recherche.

Sur le dépôt, on trouvera, en plus des données proprement dites :

— Des outils pour explorer et afficher les différentes données (sons et partitions)
— Un patch Max permettant d'expérimenter le suivi

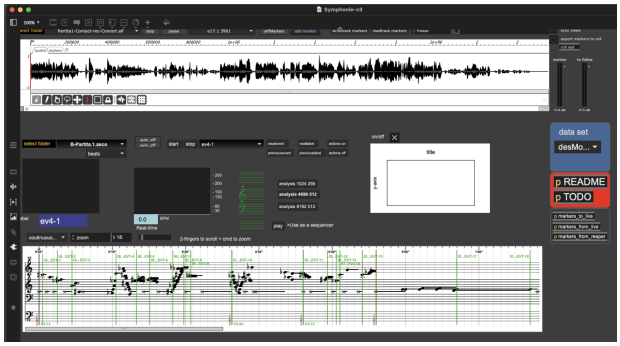

Figure 2. Visualisation du jeu de donnée "sxxi"

en temps réel avec *antescofo* des différentes pièces du corpus (Fig.3).

— Des scripts pour générer ou ré-aligner les marqueurs temporels
— Des utilitaires permettant de convertir les différents format de marqueurs provenant de logiciels variés (Wave Editor, Reaper, mubu, etc.)
— Des outils permettant d'extraire ou d'écrire les méta-données associées aux fichiers-sons (wav et aiff)
— Des scripts pour générer des évaluations off-line du suivi

5. PROSPECTIVES

5.1. Vers de nouvelles approches du suivi de partition

Dans les approches du siècle dernier, les algorithmes de suivi de partition, comme le nom l'indique, ne faisaient que "suivre" mécaniquement le musicien humain, c'est à dire que la machine n'agissait qu'en réaction. Alors que des musiciens qui interagissent entre eux sont en permanence en train d'anticiper les actions des autres, en se basant sur la notion d'un tempo commun. A fortiori, un "accompagnateur" se doit d'être toujours en avance des actions et des réactions du musicien qu'il accompagne. Sous l'influence des recherches en sciences cognitives et psychologiques liées à l'activité musicale [10], nous sommes passés du paradigme de suivi à celui d'anticipation [5]. Aujourd'hui, on cherche plutôt à modéliser le phénomène du jeu collectif par celui de synchronisation musicale entre des musiciens, qu'ils soient humains ou non, ce qui implique le partage d'une base de temps commune (le tempo) et des points de synchronisation temporelle hiérarchisés et plus ou moins précis.

Nous prévoyons de développer et de tester un nouvel algorithme de suivi de partition (reposant sur une version pre-enregistrée de reférence de la pièce suivie, tel que décrit précédemment) pouvant être utilisé avec Antescofo, dans lequel la mise en correspondance se fait en continu dans le temps, entre un flux audio en direct et un enregistrement de référence. Ce travail s'intègre dans le projet *TsimTsoum* de de l'équipe RepMus qui vise à développer une version open-source du système existant et à permettre le couplage de différentes machines d'écoute. La machine d'écoute actuelle est propriétaire : séparer *Antescofo* en composants distincts de suivi de partition et de programmation temporelle permettra la publication open source d'un vaste répertoire musical (qui, à l'heure actuelle, n'utilise que partiellement le suiveur de partition d'*Antescofo*). De nouvelles productions musicales auront ainsi le choix d'utiliser le nouveau suiveur de partition pour une réalisation entièrement open source ou de passer, au sein d'une même pièce, d'un suiveur à un autre selon le contexte musicale et les spécificités des performances de chacun. L'utilisation de logiciels open source dans les productions musicales mixtes aidera grandement les efforts de protection contre l'obsolescence constatée du répertoire de la musique mixte.

Le système résultant sera validé sur un large corpus de pièces existantes (œuvres de Philippe Manoury, Sacha Blondeau, Yan Maresz, Roger Reynolds – œuvres pour flûte, violon, alto, violoncelle, piano ou quatuor à cordes) Aucune de ces pièces ne pouvant être suivie parfaitement par la technologie existante. Puisque nous disposons de plusieurs enregistrements sonores des parties instrumentales de ces pièces, les tests peuvent être effectués sans avoir à monter une nouvelle production musicale.

5.2. Intégrer les répétitions dans le workflow du suivi

Alors qu'un musicien humain profite des répétitions collectives (c'est à cela qu'elles servent) pour améliorer sa performance, tous les environnements de suivi de partition connus jusqu'à présent sont entraînés à priori, en amont de leur première utilisation. Les nouveaux systèmes de synchronisation devront être conçus pour permettre cet apprentissage continu afin que ses performances en terme de synchronisation s'améliorent tout au long des répétitions et des exécutions successives de l'œuvre musicale. Cela implique un modèle de mémoire dynamique de toutes les exécutions précédentes. Il semble que sur ce sujet on n'ait pas dépassé le modèle initialement présenté par Barry Vercoe et Miller Puckette [19] dans lequel "il n'y a pas de 'mémoire' des interprétations, ni aucune possibilité pour le musicien synthétique d'apprendre des expériences passées. Le musicien synthétique est pour ainsi dire à chaque fois en train de déchiffrer sur scène " ("There was no performance 'memory', and no facility for the synthetic performer to learn from past experience. The synthetic performer was essentially sight reading on the concert stage every time.")

6. CONCLUSION

L'évaluation quantitative de la datation des événements musicaux constitue un prérequis méthodologique indispensable pour caractériser les performances d'un système de suivi. Des métriques objectives, telles que les taux de détection, les écarts temporels ou les mesures dérivées de type F-score, fournissent un cadre comparatif rigoureux et reproductible. Toutefois, ces indicateurs, bien qu'essen-

tiels, ne sauraient épuiser à eux seuls la question de la validité d'un système déployé en contexte réel.

En situation de concert, la pertinence d'un dispositif de suivi ne se réduit pas à la précision locale de la détection des événements. Elle dépend également de son comportement global dans la durée, de sa capacité à maintenir une cohérence musicale perceptible et de sa robustesse face aux aléas inhérents à l'interprétation vivante. Les variations expressives, les fluctuations temporelles, les imprécisions inévitables ou les ruptures momentanées du signal constituent autant de situations limites qui mettent à l'épreuve les hypothèses algorithmiques sous-jacentes. Dans ce cadre, une évaluation qualitative complémentaire s'avère nécessaire afin d'apprécier la stabilité du système, la gestion des erreurs et la musicalité résultante de son interaction avec l'interprète.

Cette évaluation élargie permet notamment d'examiner la manière dont le système absorbe ou corrige ses propres défaillances. La gestion des décrochages temporaires du suivi, l'intégration de mécanismes de rattrapage, ou encore la possibilité d'un contrôle manuel transitoire constituent des dimensions déterminantes dans un environnement scénique. Plutôt que de viser l'éradication illusoire de toute erreur, l'enjeu consiste à concevoir des architectures capables de s'accommoder de leurs limites, en maintenant une continuité fonctionnelle acceptable du point de vue musical.

Dans cette perspective, l'accès à plusieurs interprétation d'une même pièce revêt une importance particulière. Elle permet d'évaluer la sensibilité du système aux divergences d'interprétation, aux différences de tempo, d'articulation ou de dynamique, et d'observer la stabilité des performances au-delà d'un cas singulier. Une telle diversité favorise une analyse plus robuste, en évitant que les résultats ne reflètent des propriétés idiosyncratiques d'une exécution unique.

Par ailleurs, l'évolution des pratiques artistiques et des techniques de suivi invite à dépasser le seul paradigme de l'alignement audio–partition. Les dispositifs contemporains mobilisent également des alignements audio–audio ou geste–audio, ouvrant la voie à des formes d'interaction plus hétérogènes et multimodales. La constitution d'une base de données adaptée à ces différents schémas expérimentaux apparaît dès lors comme une condition nécessaire pour accompagner les développements futurs.

Le présent travail s'inscrit dans cette dynamique. Il vise à proposer un cadre structuré et documenté permettant d'évaluer des systèmes de suivi dans des conditions variées et réalistes. Il ne prétend toutefois pas établir un benchmark définitif ni universel. Les choix opérés reflètent un état des besoins et des pratiques à un moment donné, et appellent à être enrichis, discutés et adaptés à mesure que les exigences artistiques et technologiques évolueront.

7. REFERENCES

[1] Charles Ballester, Baptiste Bacot, Louis Bigo, Vanessa Nina Borsan, Louis Couturier, et al., "Interacting with Annotated and Synchronized Music Corpora on the Dezrann Web Platform", *Transactions of the International Society for Music Information Retrieval*, 2025, 8 (1), pp.121-139. ⟨10.5334/tismir.212⟩. ⟨hal-04956003⟩

[2] Bevilacqua, Frédéric, Zamborlin, B., Sypniewski, A., Schnell, N., Guedy, F. and Rasamimanana, N., "Continuous realtime gesture following and recognition", *International gesture workshop*, Springer, Berlin, Heidelberg, 2009.

[3] Caramiaux, Baptiste, Montecchio, N., Tanaka, A. and Bevilacqua, F., "Adaptive gesture recognition with variation estimation for interactive systems", *ACM Transactions on Interactive Intelligent Systems*, 4(4), 2015. `https://hal.science/hal-01266046`

[4] Cont, A., Schwarz, D., Schnell, N., et Raphael, C., "Evaluation of Real-Time Audio-To-Score Alignment", *International Society for Music Information Retrieval Conference (ISMIR)*. 2007.

[5] Cont, A. "ANTESCOFO : Anticipatory Synchronization and Control of Interactive Parameters in Computer Music", *Proceedings of International Computer Music Conference (ICMC)*, Belfast, Ireland, ICMA. 2008. `http://hdl.handle.net/2027/spo.bbp2372.2008.154`

[6] Dannenberg, R. B. "An On-Line Algorithm for Real-Time Accompaniment", *Proceedings of the 1984 International Computer Music Conference* , (pp. 193-198).San Francisco, CA : Computer Music Association. 1984. `http://hdl.handle.net/2027/spo.bbp2372.1984.025`

[7] Dorfer, M., Jr., J. H., Arzt, A., Frostel, H. and Widmer, G, "Learning Audio–Sheet Music Correspondences for Cross-Modal Retrieval and Piece Identification", *Transactions of the International Society for Music Information Retrieval*, 1(1) pp.22–33, 2018.

[8] Faghih, B., Chakraborty, S., Yaseen, A. and Timoney, J., "A New Method for Detecting Onset and Offset for Singing in Real-Time and Offline Environments". Appl. Sci. 2022, 12, 7391. `https://doi.org/10.3390/app12157391`

[9] Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., Elsen, E., Engel, J. and Eck, D. "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset". *International Conference on Learning Representations (ICLR)* 2019. `https://arxiv.org/abs/1810.12247`

[10] Huron, David; "Sweet Anticipation : Music and the Psychology of Expectation", MIT Press, 2006 10.7551/mitpress/6575.001.0001.

[11] Kong, Q., Li, B., Chen, J. and Wang, Y, “GiantMIDI-Piano : A Large-Scale MIDI Dataset for Classical Piano Music“. *Transactions of the International Society for Music Information Retrieval*, 5(1), 87–98 2022. DOI : https ://doi. org/10.5334/tismir.80

[12] Lee, L. H. “Musical Score Following and Audio Alignment”, 2022. https://arxiv.org/abs/2205.03247

[13] Music Information Retrieval Evaluation eXchange (MIREX), “Real-time Audio to Score Alignment Task”, https://music-ir.org/mirex/wiki/2006:Score_Following (consulté le 24/02/2026).

[14] Orio, N., Lemouton, S., Schwarz and Schnell, “Score Following : State of the Art and New Developments”, *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Montreal, Canada, 2003.

[15] Peter, S.D., “Online Symbolic Music Alignment With Offline Reinforcement Learning”. *Proc. of the 24th International Society for Music Information Retrieval Conference*, 634-641. 2023. https://doi.org/10.5281/zenodo.10265367

[16] Raphael, C., “Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, DOI : 10.1109/34.784467.

[17] Thickstun, J., Brennan, J. and Verma, H, “Rethinking Evaluation Methodology for Audio-to-Score Alignment”, 2020. https://arxiv.org/abs/2009.14374

[18] Vercoe, B., “The Synthetic Performer in the Context of Live Performance”, *Proceedings of the 1984 International Computer Music Conference*, (pp. 199-200). San Francisco, CA : Computer Music Association. 1984. http://hdl.handle.net/2027/spo.bbp2372.1984.026

[19] Vercoe, B., Puckette, M., ”Synthetic Rehearsal : Training the Synthetic Performer”, *Proceedings of the 1985 International Computer Music Conference*, (pp. 275-278).San Francisco, CA : Computer Music Association. 1985. http://hdl.handle.net/2027/spo.bbp2372.1985.043