Descripteurs et algorithmes de caractérisation de l'aspect rythmique du son et de la musique

Jean-Baptiste GOYEAU

DEA
Acoustique, traitement du signal et informatique appliqués à la musique

Mémoire réalisé à

IRCAM
75004 Paris
France
1 place Igor Stravinsky

Responsable : Xavier Rodet, Geoffroy Peeters

Septembre 2004

Table des matières

Remerciements

Je tiens à remercier Xavier Rodet de m'avoir accueilli au sein de l'équipe analyse-synthèse de l'IRCAM.
Geoffroy Peeters pour son encadrement et en particulier pour sa patience, sa gentillesse et la clarté de ses explications.
Benoit Meudic pour ses discussions très enrichissantes sur le rythme. Plus généralement tous ceux, dont Fabien Gouyon qui ont pu par leur travail passé éclairer mes recherches
Cyrille Defaye pour sa disponibilité et l'organisation sans faille de cette année
Tous les professeurs du DEA qui m'ont fait découvrir le coté scientifique passionnant de la musique.
Tous les étudiants de la promo 2002/2003 ATIAM pour leur sympathie et leur joie de vivre. Merci en particulier à Gaël pour les discussions matinales, à Pierre pour son aide matérielle à la réalisation de ce stage, et aux Toulousains pour le week-end montagnard.
Zora pour son soutien moral.
Mes parents pour leur soutien financier.... et tout le reste.
Karin Höthker qui m'a donné l'adresse de la maison.
Tous ceux qui ont rendu agréable et passionant mon voyage du berceau à l'IRCAM.

Résumé

Ce rapport expose le travail que j'ai effectué au sein de équipe analyse-synthèse de l'IRCAM d'avril à septembre 2004 dans le cadre du DEA ATIAM. Il montre une approche de la recherche des informations rythmiques dans un enregistrement audio. Les étapes de l'algorithme développé sont la détection d'onset et la recherche du tatum, la recherche de la pulsation à partir du tatum.

1 Introduction

L'indexation audio, que l'on peut définir comme la description réduite d'un enregistrement sonore, est aujourd'hui un domaine de recherche très actif. Parmi les objectifs poursuivis se trouvent la comparaison de plusieurs enregistrements, les recherches de documents sonores dans des bases de données très importantes en un temps raisonnable. Les informations collectées par l'indexation doivent donc être une représentation condensée de ces extraits. L´équipe Analyse-Synthèse de l'IRCAM s'est intéressée dans le cadre du projet CUIDADO à la reconnaissance d'instruments et à l´étude de paramètres pour la description des sons [Pee04]. En revanche, elle n'avait pas encore développé d'algorithme de description du rythme. Dans un premier temps, le but de l'indexation rythmique est de représenter le rythme grâce à une nombre de variables restreint. A titre d'exemple une partition comme celle de la figure 1 est une représentation possible. Dans un deuxième temps, il s'agit de fixer des mesures de similarité entre les représentations, donc des distances entre les variables des différents extraits afin de pouvoir les comparer.

Pour cela, nous devons d'abord caractériser le rythme. Nous nous baserons sur la définition du rythme par Ladzekpo [Lad89].

"Rhythm may be defined as the movement in time of individual sounds[...] however, rhythm is not only the whole feeling of movement in music, but also the dominant feature[...] [It] provides the regular pulsation or beat which is the focal point in uniting the energies of the entire community in the pursuit of their collective destiny"

L'aspect rythmique du son peut être décrit comme l'organisation temporelle de sons isolés, et plus précisément comme la répétition à intervalles réguliers de ces sons. Nous commencerons par décrire dans la partie 2 ce qu'est un son isolé, comment le détecter et le caractériser. Nous présenterons ensuite des algorithmes de recherche de l'organisation temporelle de ces sons dans la partie 3. Nous présenterons dans les parties 4, 5 et 6 les algorithmes que nous avons développés pour la représentation du rythme. Enfin, nous exposerons dans la partie 7 les résultats que nous avons obtenus.

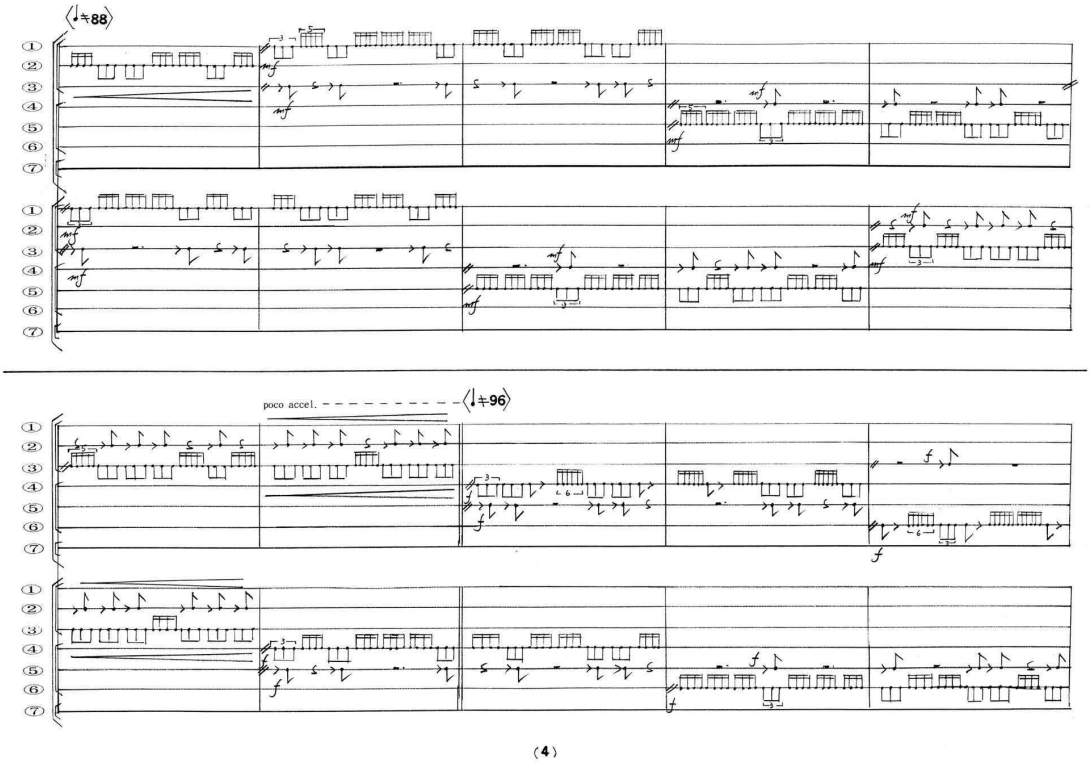

FIG. 1 – Partition d'un morceau de percussion. Extrait de Fourteen Percussions Op.119 de Maki Ishii http ://www.technogallery.com/maki-ishii/ : Les informations rythmiques globales sont la mesure et le tempo, les informations locales sont les barres de mesures, le code des longueurs des notes (blanche, noire, croche), la ligne correspondant aux notes sur la partition.

2 Caractérisation des événements sonores

Pour définir un événement sonore, commençons par donner la définition d'un objet sonore dans [Chi95] :"On appelle objet sonore tout phénomène ou événement perçu comme un ensemble, comme un tout cohérent, et entendu dans une écoute réduite qui le vise pour lui même, indépendamment de sa provenance ou de sa signification." Il y est aussi précisé le rapport entre l'objet sonore et la note de musique :"en tant qu'unité d´événement sonore qui peut-être composée de plusieurs micro-événements soudés par la forme, l'objet sonore peut ne pas coïncider, lorsqu'il s'agit d´écouter de la musique classique, avec chacune des notes de la partition : un arpège de harpe, sur la partition est un enchaînement de notes ; mais pour l'auditeur c'est un seul objet sonore".

L´événement sonore sera alors pour nous le début d'un objet sonore. En musique, un événement sera le début d'une note avec les restrictions de l'enchaînement de notes donnée dans la définition. La notion d´événement sonore correspond dans la norme midi à un onset. Un événement sonore est un atome du rythme. Nous verrons comment l'agglomération de plusieurs de ces atomes peuvent créer un rythme dans la partie 3.

2.1 Détections des événements sonores à partir du signal

Commençons par expliquer comment trouver ces événements à partir du signal et à les caractériser. L´événement sonore est caractérisé par une forte variation du signal sur instant très court. Une telle variation est appelée transitoire. Lorsqu'elle correspond effectivement à un début de note, elle est appelée transitoire d'attaque. La détection des transitoires d'attaque est un domaine de recherche en soi et a donné lieu à un grand nombre de publications. Les méthodes sont appelées méthodes de détections d'onset. L'onset est le temps du début de l'attaque. Une description complète des algorithmes est disponible dans [DS04]. Nous donnons ici un rapide aperçu de l´ensemble de ces méthodes.

Le but d'un algorithme de détection d'onsets est de repérer les débuts de notes dans un enregistrement audio d'un morceau de musique. Pour cela, on applique au signal une fonction qui donne en sortie les temps des onsets. Il procède généralement en trois étapes.

1. *Un prétraitement* : Le prétraitement consiste à projeter le signal sur une base adaptée à la fonction de détection. Les pré traitements les plus utilisés sont la séparation sinusoïde/bruit (par exemple.[ABDR03]) ou les séparations par bandes de fréquences ([Sch98],[Kla99]). La séparation sinusoïde-bruit est intéressante car les parties bruitées sont très marquées lors des transitoires.

   La séparation en bandes de fréquences est liée à notre perception du son : le son qui arrive dans l'oreille est transformé en un signal propagé

dans la membrane basilaire. Selon sa hauteur spectrale, il va exciter des parties différentes de cette membrane et un signal différent va être transmis au cerveau. Scheirer a modélisé ces différentes excitations en observant non pas le signal dans sa totalité mais par bandes de fréquences (correspondant à chaque excitation-environ 1/3 d'octave).

2. *L'application d'une fonction de détection au signal pré traité* : C'est le point critique de la détection. La fonction de détection doit mettre en évidence l'apparition des transitoires d'attaques. Ces transitoires sont repérables grâce à certaines caractéristiques du signal. Le signal est sous-échantillonné pour pouvoir les calculer. Les principales méthodes sont :

   (a) *l'observation des caractéristiques temporelles.* : Les premières méthodes observent l'enveloppe d'énergie du signal [Sch98]. Il s'agit d'un filtrage passe-bas de l´énergie du signal. En effet, pour certains signaux (dont les signaux percussifs), les transitoires sont caractérisés par une forte augmentation de l´énergie. Ces méthodes ont ensuite été améliorées. D'une part grâce aux prétraitements décrits plus hauts (séparation sinusoïde/bruit, séparation par bandes de fréquence). D'autre part grâce à l'observation, non plus de l´énergie mais de la dérivée du logarithme de l´énergie [Kla99]. Cette méthode met en évidence l'importance de la perception de la variation du volume sonore : la dérivée du logarithme de l´énergie est la variation du volume en décibel.

   (b) *l'observation des caractéristiques spectrales.* Ces méthodes utilisent la transformée de Fourier à court terme (TFCT). Les transitoires d'attaque sont caractérisés par une forte variation des coefficients complexe de la TFCT. La fonction observe soit l'augmentation conséquente de l'amplitude de certains coefficients[JR01], soit une rupture dans l'évolution de la phase. La méthode sur la phase est sensible au bruit.

   (c) *la rupture de modèle[BN93].* Ces méthodes sont statistiques. On considère chaque note comme un modèle statistique. Il y a deux approches différentes. La première considère que toutes les notes ont le même modèle et que les transitoires se situent au moment où le signal s´éloigne du modèle. L'autre considère deux modèles qui alternent au moment de la transition entre deux notes. Ces méthodes sont efficaces mais complexes.

Nous nous sommes contentés de la première approche (a), car nous avons essentiellement observé des signaux à forte variation d'énergie. Elle nous a donné des résultats satisfaisants (cf paragraphe 7.2). Nous détaillerons dans le paragraphe 4.2.2, la fonction de détection de Klapuri [Kla99], dont nous nous sommes inspirés.

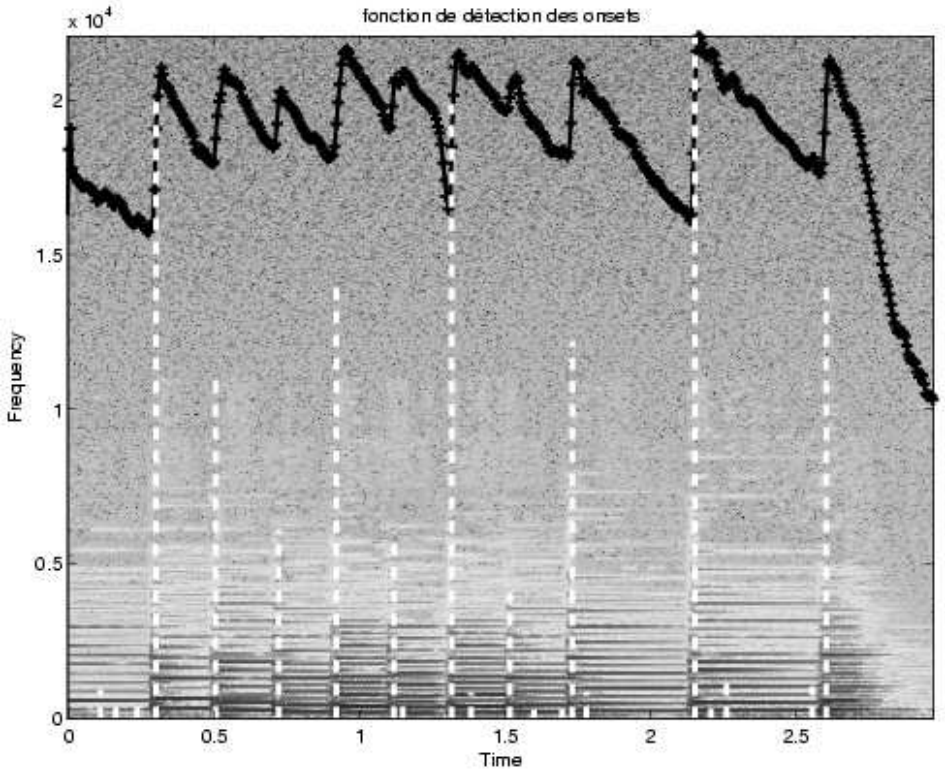

FIG. 2 – Exemple de fonction de détection pour un enregistrement de piano. La fonction de détection est tracée en noir sur le spectrogramme. Les onsets détectés sont représentés en pointillés blancs. La hauteur des pointillés blancs représente l'importance de la variation de la fonction de détection au moment où l'onset a été détecté. NB : La fonction n'est pas à l'échelle. Elle est tracée de façon à avoir son maximum en haut du spectrogramme

3. *La recherche de maxima pertinents de cette fonction* : Une fonction de détection est telle que les onsets se situent aux maxima de cette fonction. Pourtant certains maxima ne sont pas pertinents. Par exemple, deux maxima proches l'un de l'autre peuvent être liés à un même onset. D'autres maxima peuvent être lié à une variation de bruit de fond. Pour éliminer ces onsets parasites, une méthode souvent employée est d'appliquer un seuil adaptatif à la fonction de détection en dessous duquel les maxima ne seront pas considérés comme onsets.

La figure 2 montre un exemple de fonction de détection avec la position des onsets détectés pour un enregistrement de piano.

La figure 3 présente la structure "type" d'un algorithme de détection d'onsets.

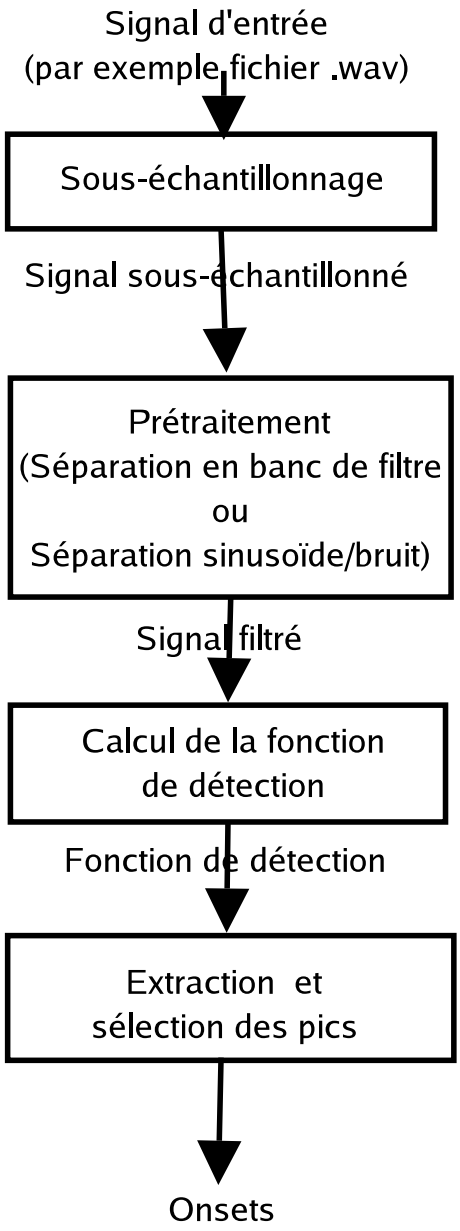

FIG. 3 – Algorithme "type" de détection d'onset.

2.2 Caractérisation d'un événement : notion de timbre

Après avoir détecté les événements il s'agit de les caractériser. L'une des caractérisation possible est leur amplitude, ou plutôt leur perceptibilité. Non seulement, on doit considérer avec un poids plus fort les événements que l'on entend mieux, mais leur dynamique joue une rôle dans la perception du rythme. Nous le verrons dans les chapitres suivants.

L'autre caractéristique perceptivement importante est le timbre. Chaque instrument joue une partie différente qui a un rythme différent. Nous percevons chacun de ces rythmes. Certaines études tentent de classer les événements selon leur timbre [HDG03]. Sans aller jusqu'à la classification de ces événements, nous avons tenté de caractériser des onsets suivant certains paramètres de signal pour établir des critères de similarité.

3 Analyse de l'organisation temporelle

Nous avons décrit dans la partie 2 les atomes du rythme que sont les événements sonores. Le rythme est l'organisation temporelle de la succession de ces événements. Nous montrons dans la partie 3.1 l'importance perceptive des périodicités associées à ces événements. Nous définissons ensuite dans la partie 3.2 les différentes périodicités utiles à la description du rythme.

3.1 Description de la perception du rythme

Lorsque le rythme est régulier, nous pouvons le décrire en terme de périodicité de répétition. Ces périodicités sont dans une gamme de valeur beaucoup plus faible que les fréquences caractérisant la hauteur des notes. Pour le rythme, les périodicités étudiées sont généralement comprises entre $\frac{40}{60}$ Hz et $\frac{200}{60}$ Hz, alors que la hauteur d'un premier partiel peut aller de 40 à 2000 Hz. Pour cette raison, les périodicités relatives au rythme sont exprimées dans une unité spécifique, le battement par minute. Un battement par minute est égal à $\frac{1}{60}$ de Hz. Cette unité sera notée dans la suite bpm. Les propriétés perceptives que nous décrirons sont tirées de [P F74].

3.1.1 Le battement, manifestation principale de la perception du rythme

Définition de la pulsation, du tactus et du battement L´écoute d'un morceau engendre souvent chez l'auditeur un mouvement régulier du pied ou un balancement du corps. Ces mouvements seront appelés dans la suite *battements* (traduction française du beat anglais). L'intervalle entre deux battements est appelé *pulsation* et est exprimé en millisecondes. Cette pulsation peut varier au cours du temps. Par exemple, elle accélère lorsque le tempo accélère. La périodicité des battements,

appelée *tactus*, se situe généralement entre 40 bpm et 200 bpm [P F74]. Le paragraphe 3.1.2 tente de préciser les phénomènes liés au battement.

3.1.2 Importance du groupement perceptif : définition des patterns rythmiques

Remarquons d'abord que les battements sont synchrones avec les événements rythmiques décrits dans la partie 2. Ce sont les événements qui crée le rythme, mais pas seulement. L'organisation de ces événements est très importante. La perception du rythme commence lorsqu'une succession d´événements est perçue comme un tout. On peut citer par exemple les trois coup d'une horloge. La tendance à regrouper des phénomènes individuels pour former un tout est étudié sous le nom de Gestalt. On peut énumérer quelques lois perceptives pour la création d'un groupe rythmique :

1. Lorsque les événements sont réguliers - par exemple, la chute de gouttes d'eau d'un robinet mal fermé -, on les perçoit par groupe de 2 ou 3, quelquefois 4. Les regroupements pour des groupes plus importants sont en fait des regroupements successifs. Par exemple, un groupement par cinq est un groupement par deux puis par un groupement par trois.
2. L'intervalle entre deux groupements subjectifs apparaît plus long que les intervalles entre les éléments du groupe.
3. Le premier événement du groupe (et parfois le dernier) apparaît plus accentué que les autres.
4. La succession des événements provoque un oubli : les structures reconnues sont centrées au moment de l'écoute.
5. Les limites fréquentielles du groupement sont du même ordre que celles observées pour les battements. En dessous de 30 bpm, les événements de chaque groupe sont perçus individuellement, au dessus de 300 bpm, la structure perçue est le groupement de plusieurs structures.

Toutes ses observations sont utilisées implicitement dans la musique occidentale. Par exemple, on utilise la perception d'un temps fort sur le premier élément d'un groupement : le premier temps de la mesure est fort et les suivants sont faibles. Ainsi, le groupement de la mesure est perçu plus facilement par un auditeur.

La Gestalt révèle deux aspects importants de la perception du rythme. Premièrement, elle explique la constitution de groupes perceptifs de plusieurs événements. Mais elle explique aussi l'importance de la répétition de ces groupes. Un morceau est une succession de formes rythmiques. L'apparition de structures rythmiques est due à un retour plus ou moins périodique de groupements identiques ou analogues. Il peut tout de même y avoir des variations autant dans la durée des structures que dans leur constitution.

C'est la succession de structures isochrones qui est à l'origine du battement, et donc du rythme.

L'analyse du rythme doit prendre en compte ces deux aspects : La constitution d'un ou plusieurs groupes d'événements et la succession isochrone de ces groupes.

3.2 Constitution d'une hiérarchie de fréquences perceptives

La succession isochrone de structures rythmiques définit périodicité unique qui est l'inverse la durée d'une de ces structures. Toutefois, dans la grande majorité des morceaux de musique, ces structures sont construites de façon à renforcer la perception de cette périodicité : les écarts temporels entre les événements constitutifs d'une structure rythmique sont souvent des sous-multiples de la durée de la structure. D'autre part, la succession des structures forment des structures de durées plus longues, elles aussi répétées. Ainsi, un morceau est constitué d'une hiérarchie d'intervalles multiples les uns des autres, dont certains comme la pulsation correspondent à la durée de structures répétitives d'un morceau. La figure 4 présente l'ensemble des des termes que nous définissons dans ce paragraphe.

L´écriture d'un morceau de musique est d'ailleurs basée sur cette hiérarchie. Sur une partition, la durée des notes est déterminée à partir d'une valeur de référence. Dans la notation occidentale cette valeur de référence est notée la plupart du temps par une noire, une blanche ou une croche. La notation permet ensuite de représenter des notes de durées multiples ou diviseurs de la durée de référence.

Définition des intervalles inter-onsets(IOI) Il y a donc plusieurs intervalles multiples les uns des autres entre les différentes notes. On appellera dans la suite intervalle inter-onset (abrégé en *IOI*) la durée entre deux notes ou onsets non forcément successifs.[1].

Certains intervalles ont alors dans cette hiérarchie une valeur particulière.

Définition de la micropulsation et du tatum La *micropulsation* est l'intervalle le plus petit entre deux notes successives. La fréquence associée à cet intervalle s'appelle le *tatum*. Le tatum a été introduit dans [Bil93] comme la fréquence associée à la pulsation la plus rapide que l'on perçoit dans un morceau. Il est précisé que c´est une subdivision du tempo. Si le morceau est parfaitement régulier, c'est donc l'intervalle de temps minimum séparant 2 onsets ou plus précisément le plus grand commun diviseur des intervalles inter-onsets (non nécessairement successifs). Lorsque l'on bat la micropulsation, tous les onsets sont sur les temps, et il n'existe pas d'intervalle de temps plus petit vérifiant cette

[1]Les recherches sur la reconnaissance du rythme ont montré l'importance de considérer également les intervalles de temps entre les onsets non successifs [GHC02]

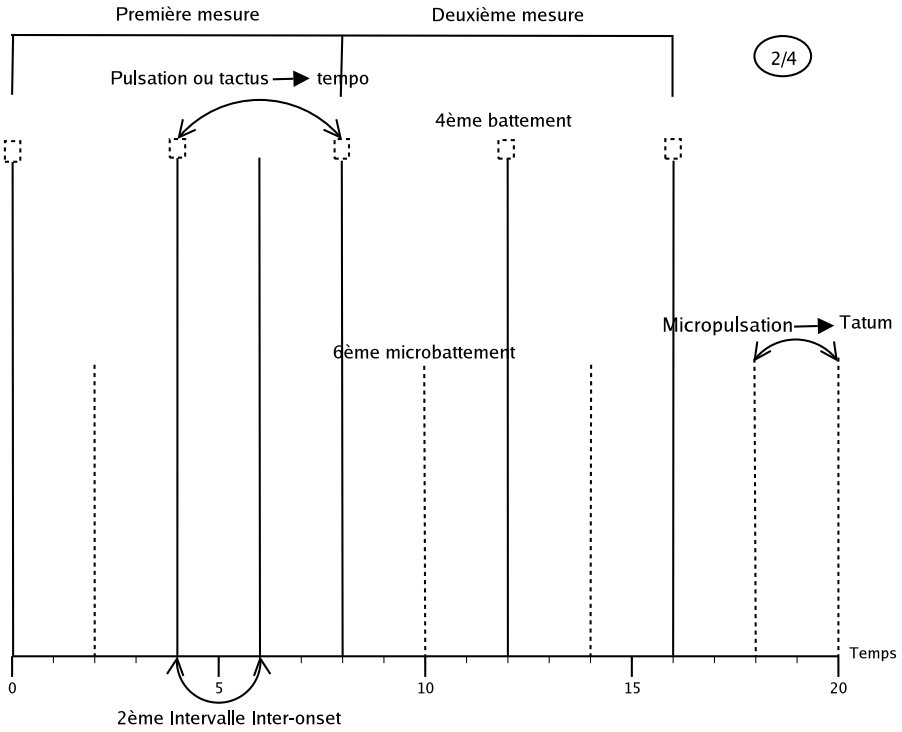

FIG. 4 – Présentation des termes relatifs au rythme : les onsets sont les barres noires, les battements sont les carrés au-dessus des barres, l'exemple est une mesure à deux temps.

propriété. Les *intervalles interonsets* d'un morceau sont des *multiples de la micropulsation*. On peut donc construire une grille de *micro-battements* espacés d'une micropulsation. Sur cette grille, on pourra placer tous les onsets. C'est l'intérêt principal de la micropulsation : elle permet de quantifier les IOI.

Définition du tempo Le *tempo* définit la durée de la valeur de référence. On déduit les valeurs des autres durées de notes par multiplication ou division. Le tempo définit la rapidité de l'exécution d'un morceau. La Marseillaise a par exemple un tempo de 120 bpm à la noire. Il y a 120 noires dans une minute. On peut souvent la rapprocher du battement. On peut imaginer que le battement principal du compositeur est le tempo du morceau.

Définition de la mesure Les notes sont aussi regroupées en structures de haut niveau. Les plus importantes sont les *mesures*. Ce sont des structures de durée égale. Elles sont souvent marquées dans la musique occidentale par l'accentuation de leur première note. Ces mesures peuvent ensuite aussi être regroupées en entité de plus haut niveau. Chaque entité d'un même niveau a une durée identique. Les différentes entités ont souvent des caractéristiques communes. La recherche de ces caractéristiques peut permettre de retrouver les structures.

3.3 Algorithmes de détection du tatum et du tactus

3.3.1 Algorithme de détection du tatum

Dans [Sep01], le tatum est définit comme le plus grand commun diviseur approché de l'ensemble des IOI, comme on l'a fait remarqué dans la partie 3.2.

Une façon plus musicale de voir le tatum est la suivante : le tatum est équivalent à la fréquence fondamentale du rythme : c'est le plus grand intervalle de temps dont les multiples expliquent les interonsets du morceau, tout comme la fréquence fondamentale est la plus grande fréquence dont les multiples expliquent les harmoniques d'un son. Les algorithmes de détection du tatum se rapprochent donc des algorithmes de fréquence fondamentale.

Dans [GHC02], Fabien Gouyon présente un exemple de caractérisation rythmique sur des fichiers audios de percussions. Après avoir détecté les onsets par une méthode basée sur l'augmentation de l´énergie, il essaie de retrouver le tatum. Il utilise pour cela les interonsets. Sa méthode consiste à tracer l'histogramme des interonsets et à mettre en évidence le PGCD des interonsets grâce à un algorithme de détection de fréquence fondamentale. L'algorithme utilisé par Fabien Gouyon est le Two-way mismatch error function [J.B93]. L'originalité de son travail est d'avoir mis en valeur l'intérêt de la définition de l'interonset multiple donnée dans le paragraphe 3.2.

L'intérêt du tatum est que sa définition prête moins à confusion que la définition de la pulsation. La pulsation en effet vit au dépens de celui qui l´écoute. Le prochain paragraphe montre cette incertitude pour les algorithmes de recherche du battement (tactus).

3.3.2 Algorithme de détection du tactus/battement

Dans la littérature, le battement est la périodicité du rythme la plus étudiée. Cela est sûrement dû à son importance perceptive(cf 3.1.1).

Remarques préliminaires : L'analyse d'un morceau de musique ne peut déterminer infailliblement une pulsation. Certains multiples ou diviseurs de la pulsation peuvent eux-mêmes être considérés comme pulsation. A l'écoute, des personnes différentes peuvent estimer des pulsations différentes pour un même morceau. Pour modéliser cette non unicité, on peut établir une liste de pulsations possibles [Meu04]. D'autre part, la plupart des algorithmes recherchent la pulsation dans les plages de fréquences entre 60 et 200 bpm (ou autour de 60 bpm), pour intégrer les propriétés perceptives. Certains considèrent aussi comme correct de trouver une pulsation multiple de la pulsation indexée. L'évaluation d'un algorithme d'estimation du tempo, n'est pas aisé.

Les algorithmes de détection du battement peuvent prendre plusieurs formes :

1. Algorithmes basés sur l´écart temporel entre les onsets

   Certains algorithmes se basent essentiellement sur la position temporelle des onsets. Ils considèrent que la pulsation est l'IOI le plus présent dans la séquence [GHC02].

   Une amélioration apportée par [Meu04] est de donner un poids perceptif à chaque onset : plus ce poids est important, et plus une pulsation possible qui tomberait sur cet onset prend de la valeur. Les paramètres qui influencent le poids pour la pulsation sont :

   (a) la répétition d'interonset : si un intervalle interonset est répété, l'onset détecté au milieu est perceptivement mieux entendu.

   (b) la durée : plus un interonset est long, plus on entend l'onset situé au début de l'IOI. On peut rapprocher cette accentuation de l'accentuation des pauses (paragraphe 3.1.2)

   (c) la densité des accords : plus un onset contient de notes, mieux il est perçu.

   (d) les fins d'onsets : on entend plus un onset qui correspond à la fin d'une autre note.

   Cette liste est heuristique et a été testée sur une base de données ; toutefois on peut imaginer donner des règles différentes. L'avantage de cette méthode entièrement temporelle est qu'elle donne aussi la phase

des battements : Elle donne non seulement la pulsation mais aussi la position temporelle des battements.

2. Algorithmes basés sur l'analyse par bandes de fréquences
   L'algorithme de référence est celui de Scheirer [Sch98]. La déduction du battement se fait à partir la variation de l'énergie. Pour cela des résonateurs sont utilisés dans chaque bande de fréquence. On cherche pour chaque bande de fréquence, le résonateur d'énergie la plus présente. La périodicité de répétition est alors celle de ce résonateur. Le tempo est la périodicité la plus présente dans l'ensemble des bandes. La recherche par bandes de fréquence est aussi utilisée dans [ABDR03]. Les onsets sont déterminés dans chaque bande de fréquences. La pulsation est le maximum de la somme des fonctions d'autocorrélation des onsets par bande de fréquence.
3. Algorithmes basés sur la caractérisation des onsets
   Une autre méthode utilise une fonction d'autocorrélation sur des événements précis. La première étape consiste à détecter le tatum (cf 3.3.1). Ensuite, on recherche les diviseurs pertinents du tatum. Ce sont ceux pour lesquels la fonction d'autocorrélation (de paramètres pertinents perceptivement du signal) est maximale. Par exemple, on peut classer les onsets par catégories (cf 2.2) et étudier les écarts temporels entre chaque élément d'une même catégorie.
   Remarque : la méthode de [ABDR03] utilise comme catégorisation les bandes de fréquences.

3.4 Variabilité du tempo

On trouve rarement un tempo ou un tatum unique pour l'ensemble du morceau. On ne peut donc intégrer les algorithmes de détection du tempo sur l'ensemble d'un morceau. Nous estimons le tempo trame par trame sur des trames de durée égale à trois secondes

3.5 Définition d'un IOI-gramme

Nous avons vu dans les paragraphes précédents que la répartition des intervalles interonsets est une représentation possible des périodicités du rythme. Par exemple, on peut définir le tatum comme PGCD de ces intervalles. Il existe aussi des algorithmes qui prennent le maximum de l'histogramme des IOI comme définition de la pulsation [GHC02]. Si on déroule cet histogramme au cours du temps en prenant des fenêtres de 3 secondes pour analyser les intervalles interonsets, on obtient un IOI-gramme des intervalles temporels représentatifs du morceau. Un exemple est donné par la figure 5. La micropulsation a une place privilégiée puisque c'est le plus petit intervalle. On obtient ensuite un répartition des IOI sur les multiples de cette

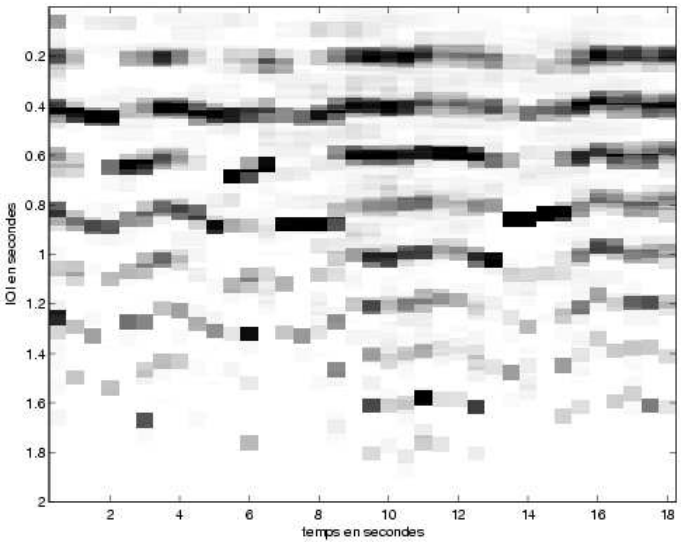

FIG. 5 – IOI-gramme pour un extrait de Weeping-Willow de Scott Joplin : Les lignes horizontales noires représentent les IOI très marqués. La ligne supérieure est celle de la micropulsation (croche). Le morceau est à 2/4. Aux alentours de la huitième et de la 14ème seconde l'interonset correspondant à la blanche, unité de la mesure, est accentué

micropulsation. Parmi ces multiples se trouvent la pulsation et les différents niveau de la métrique. On obtient donc un spectre de raies correspondant aux périodicités du rythme. Il s'agira alors de pondérer ces intervalles par leur importance perceptive pour obtenir un suivi de périodicité rythmique. D'après ce que nous avons vu, cette périodicité est locale, ce qui est le cas de notre représentation. D'autre part, elle est basée sur l'accentuation, nous verrons comment cela est pris en compte dans la représentation. Nous essaierons dans la suite de montrer l'importance de ce diagramme et de montrer les algorithmes qui permettent de donner un poids perceptif à chaque intervalle inter-onset.

4 Algorithme de détection des onsets

4.1 Représentation des onsets

Le point de départ de notre étude est de définir une représentation probabiliste des onsets : à chaque onset $Onset_i$ détecté pour l´échantillon n_i, est associé un *indice de confiance* $P_{Onset}(n_i)$. Une fonction de vraisemblance est ainsi définie : À un échantillon n, elle associe la probabilité $P_{Onset}(n)$ qu'il y ait un onset à cet instant. Cette fonction est nulle aux instants où un onset n'a pas été détecté. $P_{Onset}(n)$ doit finalement représenter la probabilité

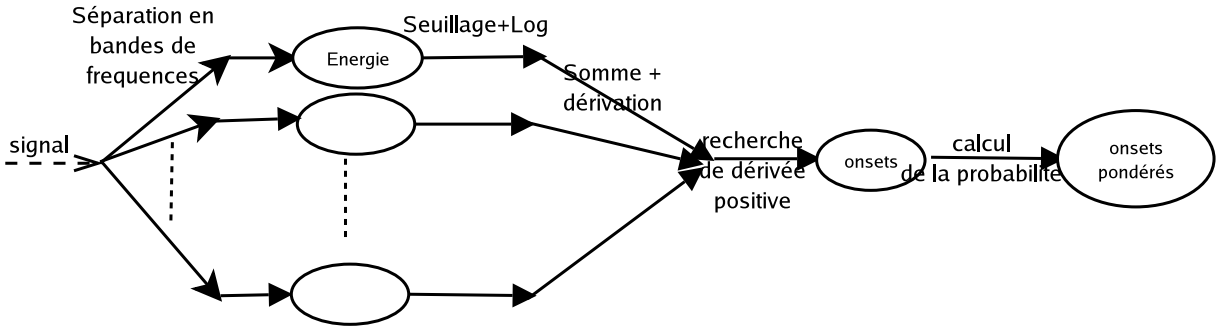

FIG. 6 – Présentation de notre algorithme de détection d'onsets

d'entendre un onset à l'instant n.

Remarque 1 : La notion de probabilité est à rapprocher de la notion de poids pour trouver la mesure définie dans [Meu04]. Dans [Meu04], la méthode s'appuie sur les onsets forts pour marquer les temps forts. Nous nous appuyons sur les onsets qui ont une forte probabilité pour trouver les informations rythmiques. Malgré tout, il y a une différence fondamentale entre cette approche et la nôtre. Les onsets considérés dans [Meu04] sont ceux d'un partition et auraient donc tous une probabilité de 1 dans une représentation probabiliste. La notion de probabilité indique que notre détection n'est pas infaillible et que les onsets de faible probabilité peuvent être de fausses détections. Nous montrerons comment l'influence de ces faux onsets peut être réduite après la recherche du tatum. Nous supposons que notre détection est assez fiable pour que ces onsets n'influencent pas la recherche du tatum.

Remarque 2 : Un onset $Onset_i$ est détecté lors du dépassement d'un seuil définit par la *fonction de détection* f_{Onset}, fonction du temps. Dans [Kla99], par exemple, le seuil est un pourcentage de l´énergie maximale. Si le seuil est bien construit, plus $f_{Onset}(Onset_i)$ dépasse le seuil, plus la probabilité que cet onset soit entendu par un auditeur est forte. La notion de probabilité dépend donc de la méthode de détection des onsets (en particulier de la fonction de détection). Nous verrons ce que cela implique dans la partie 4.2.3. Nous utiliserons aussi dans la suite la *fonction de présence d'un onset* $F_{presence\ onset}$. $F_{presence\ onset}(n)$=1 si un onset est détecté à l'instant n, 0 sinon.

4.2 Présentation de l'algorithme de détection des onsets

Dans [KP02], A. Klapuri donne une méthode de détection des onsets basée sur la variation de l'énergie. Notre méthode se rapproche fortement de la sienne :

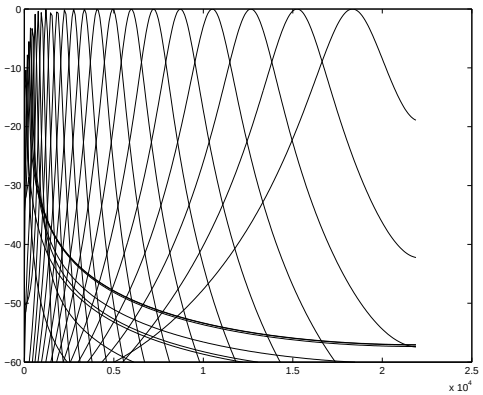

FIG. 7 – Filtres de ERB pour une division en 24 filtres d'un échantillonnage à 44100 Hz

4.2.1 Découpage en bandes de fréquence

Klapuri rappelle que les systèmes de détection du rythme utilisent un découpage en banc de filtre (l'explication est donnée au paragraphe 2.1). Celui que nous utilisons est un banc de filtre de ERB (Equivalent Rectangular Bandwidth - cf [MG83]). Ce filtrage représente la largeur de bande critique de masquage pour une fréquence donnée. Ainsi nous aurons pour cette bande de fréquence le signal reçu par la membrane basilaire, tel qu'on présume qu'elle le reçoit. La figure 7 représente une série de filtres de ERB.

4.2.2 Détection d'un onset

Un onset induit une variation du signal. Klapuri indique que l'on peut traduire cette variation par une augmentation de l´énergie dans une ou plusieurs bandes de fréquence, en première approximation. Cette approximation est bonne pour les onsets percussifs. Par ailleurs, il indique que c'est la variation relative du niveau d'énergie qu'il faut observer, donc la variation en décibel. Ce modèle rejoint le modèle de perception d'un son par l'oreille. Il cherche donc les maxima de la fonction f_{onset} définie pour chaque bande de fréquence b par l'équation 1 :

$$f^b_{onset}(b,n) = \frac{\partial(Log(Ener(b,n))}{\partial n} \tag{1}$$

La fonction effective utilisée dans [KP02] est :

$$f^b_{onset}(b,n) = \frac{\partial Log(1 + J * Ener(b,n))}{\partial n} \tag{2}$$

où J est un facteur égal à 1000. Le facteur 1000 est indiqué comme n´étant pas critique. Cela permet d'observer une courbe positive et de compresser

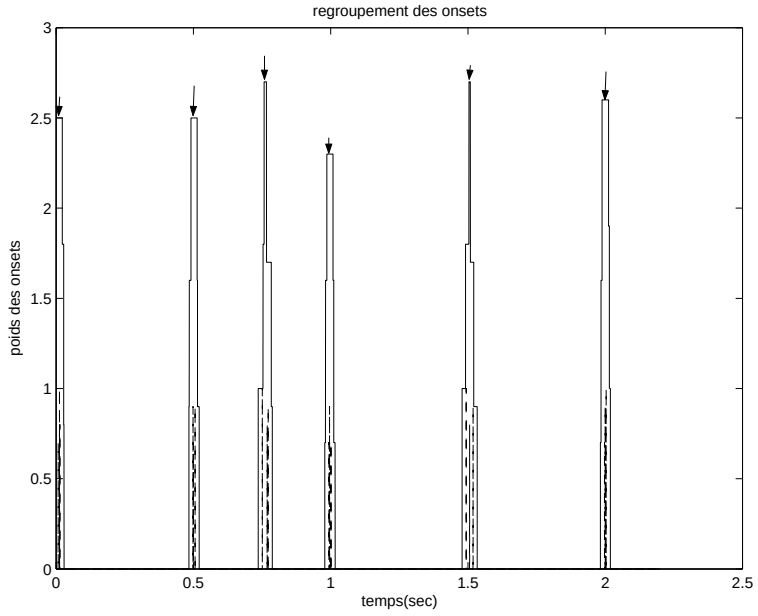

FIG. 8 – Exemple de regroupement d'onsets. Les onsets détectés par bandes de fréquences sont en pointillés, la convolution est en trait plein et les onsets finaux retenus sont marqués d'une flèche

le signal (afin d'avoir des variations plus homogènes sur une longue durée). Nous noterons dans la suite : $Ener_{Klapuri}(b,n) = Log(1 + J * Ener(b,n))$

Nous utilisons la même fonction de détection. Afin de limiter l'influence des variations d´énergie à bas niveau sonore (bruit de fond), nous considérons qu'en dessous d'un seuil d´énergie, l'énergie est constante et égale à ce seuil (donc sans variations).

Klapuri détecte les pics $Onset_i^b$ de la fonction de détection $f_{onset}^b(b,n)$ dans chaque bande de fréquence. Il somme ensuite les onsets proches temporellement (convolution par une fenêtre de 50 ms) pour trouver les onsets finaux (voir figure 8). Il conserve ensuite les onsets dont la valeur dépassent un certain seuil :

$$F_{presence\ onset} = \sum_{b \in bandes} Onset_i^b \tag{3}$$

où $F_{presence\ onset}$ est la fonction de présence des onsets. Nous avons obtenu de meilleurs résultats en faisant la somme des fonctions de détection sur l'ensemble des bandes et en recherchant les onsets comme maxima de la fonction de détection globale.

$$F_{presence\ onset}(n) = \text{maxima}(\sum_{b \in bandes} f_{onset}^b(n)) \tag{4}$$

Cette méthode est également employée dans [URCH04]. L'avantage est que le seuillage est fait sur l'ensemble du signal. En effet, certains onsets ne sont pas détectés simultanément dans toutes les bandes de fréquences (par exemple les hautes fréquences d'un onset de piano sont en retard par rapport aux basses). Le regroupement temporel des onsets détectés par bandes ne nous a pas semblé trivial en raison de ces différences temporelles entre bandes.

Notre fonction de détection est donc :

$$\sum_{bandes} \frac{\partial Ener_{Klapuri}}{\partial n}. \tag{5}$$

Nous donnerons dans la suite le nom de *fonction d´énergie* la fonction :

$$F_{energie} = \sum_{b \in bandes} Ener_{Klapuri}(b, n). \tag{6}$$

Nous considérons uniquement les dérivées positives : nous considérons comme onset toute augmentation de la fonction d´énergie.

4.2.3 Choix de la définition d'une probabilité

Comme nous l'avons vu dans la remarque 2 du paragraphe 4.1, l'importance de l'onset est liée à la fonction de détection. Plus la valeur de la fonction de détection est forte, plus la probabilité d'avoir un onset est grande. Notre fonction de détection est la variation de la fonction d´énergie. Nous pouvons prendre cette fonction de détection (pente du logarithme de l'énergie) comme départ pour estimer la probabilité. Toutefois, il semble plus intéressant de prendre la différence entre la maximum et le minimum de la fonction d´énergie autour de l'onset. On a ainsi la différence de décibel entre le son avant l'onset et le maximum de l´énergie de l'onset. La figure 9 montre l'utilité d'utiliser la fonction d´énergie directement et non sa dérivée.

L'importance d'un onset est donc proportionnelle à $max(F_{energie}(Onset)) - min(F_{energie}(Onset))$(Voir figure 10.)

Enfin, la probabilité d'un onset dépend le contexte. Nous considérons que la probabilité d'un onset est la variation locale de l'énergie autour de l'onset divisée par le maximum des variations des énergies des autres onsets. C'est grâce à cet algorithme que nous avons obtenu la figure 2.

4.2.4 Choix de la fenêtre pour le calcul de l'énergie : attribution d'une position à l'onset

Le calcul de l´énergie est fait sur une fenêtre de 30 ms. Le pas d'avancement est de 20% de la longueur de cette fenêtre. Notre sous-échantillonnage est donc à 6ms. Nous avons effectué des tests sur la base de données de sons percussifs que nous détaillons dans la partie 7.1. La performance de

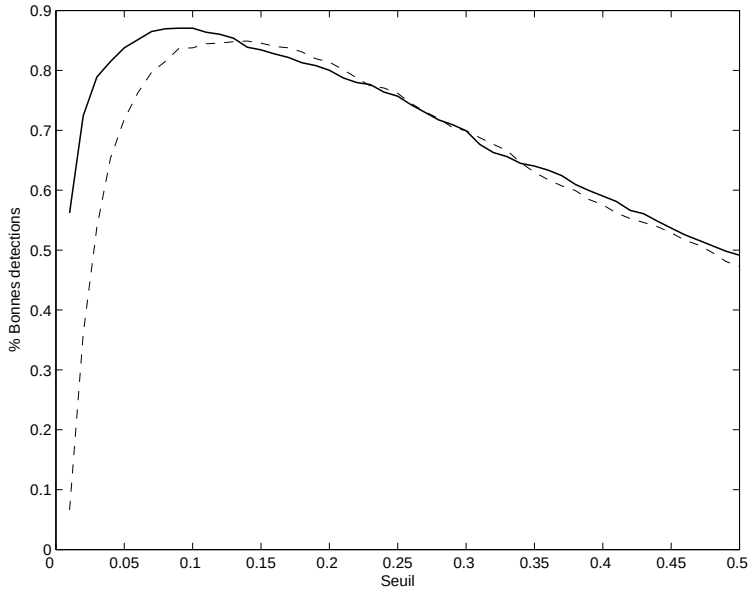

FIG. 9 – Comparaison des fonctions de détection sur la base de données des onsets décrite dans le paragraphe 7.2 : les courbes représentent le pourcentage de bonnes détections en fonction du seuil. Plus la courbe est haute, meilleure est la fonction de détection. La courbe en trait plein est celle obtenue en utilisant la fonction d´énergie entre le début et la fin de l'onset, celle en pointillés par la fonction de détection. C'est surtout pour les faibles seuils (à gauche) que la différence est importante

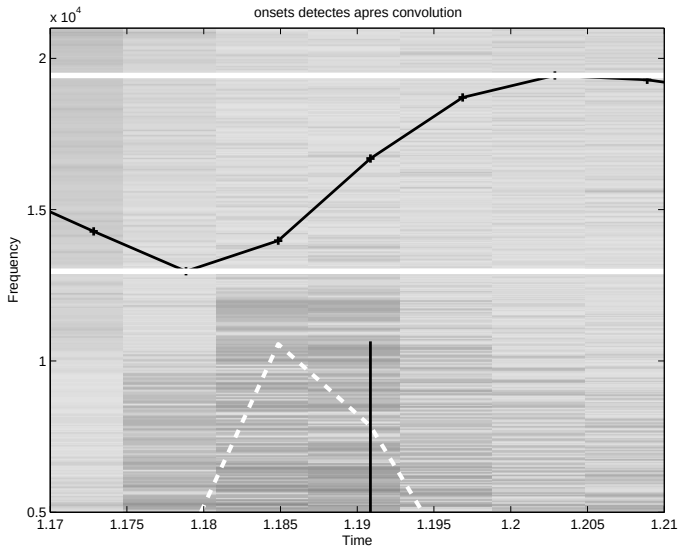

FIG. 10 – La fonction d'énergie est en noir, la fonction de détection est en blanc pointillé. L'importance de l'onset est la différence entre les deux traits horizontaux correspondants au minimum et au maximum de l´énergie autour de l'onset

fenêtre	écart moyen estimation/observé	variance de l'écart
Hanning	5ms	5ms
1/2 Hanning	11 ms	5ms

TAB. 1 – Comparaison des écarts observés pour l'estimation de la position d'un onset suivant 2 types de fenêtres

l'algorithme ne dépend pas de la forme de la fenêtre. En revanche, la position temporelle de l'attaque en dépend. Nous estimons le début d'un onset en observant l'augmentation brusque de l'enveloppe du signal lorsque c'est possible. Pour ces onsets nous comparons la position marquée à la main avec celle de l'algorithme. Nous obtenons le meilleur résultat en utilisant une fenêtre de Hanning et en positionnant l'onset au milieu de la fenêtre pour laquelle la fonction de détection $f_{onset}(n)$ est maximale. Nous obtenons ainsi une erreur moyenne de 5ms environ et une variance de l'erreur de 5ms environ. Si on prend une demi-fenêtre de Hanning, on obtient le meilleur résultat en positionnant l'onset à la fin de la fenêtre dans laquelle la fonction de détection devient positive, mais l'erreur moyenne est plus grande (11ms environ).

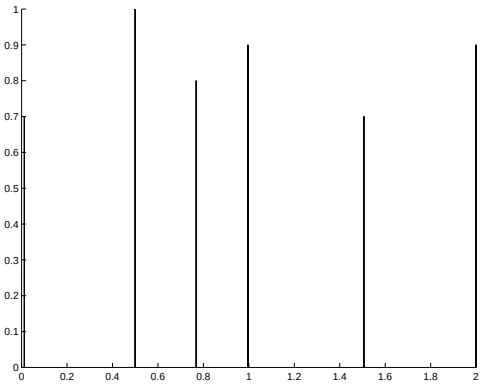

FIG. 11 – Exemple de détection d'onset obtenue sur les deux premières secondes d´un morceau de percussion. En abscisse le temps, en ordonnée la probabilité de chaque onset. Nous nous servirons de cet exemple pour illustrer notre algorithme dans les figures suivantes.

5 Algorithme d'estimation du tatum

Nous illustrons cette partie avec un exemple de détection d'onsets représenté sur la figure 11

5.1 Nécessité d'une approche par trame

Pour détecter les informations de pulsation ou de micropulsation, [Meu04] utilise une méthode propagative. Il modifie la pulsation en fonction de l'onset qu'il découvre. Dans [Meu04], l'onset est observé , dans notre cas, nous n'avons qu'une probabilité d'observation. En cas de fausse détection ou de non détection, de fausses pulsations induites peuvent se propager. Nous ne pouvons donc nous baser sur son algorithme. Afin de rendre l'algorithme plus robuste, nous utilisons l'information de l'ensemble des probabilités des onsets sur une fenêtre de quelques secondes. Ainsi, nous espérons que les fausses détections auront moins de poids. A ce titre, l'algorithme d'estimation du tatum de Fabien Gouyon [GHC02] paraît le plus adapté.

5.2 Représentation des intervalles interonsets

Un intervalle interonset ioi_{n_1,n_2} est la distance temporelle entre 2 onsets $Onset_{n_1}$ et $Onset_{n_2}$. Nous souhaitons de plus donner un poids à chaque interonset. Ce poids définit l'importance de l'interonset. Il faut remarquer que cette probabilité définit la probabilité d'existence d'un interonset entre 2 positions temporelles particulières. Ce n'est pas une a prorement parlé une

probabilité puisque :

$$\sum_{n_1,n_2\in[1,N]^2} P_{ioi}(n1,n2) \neq 1 \tag{7}$$

$Onset_{n_1}$ et $Onset_{n_2}$ ont des importances indépendantes et donc des probabilités indépendantes. On peut donc prendre comme poids de l'interonset le produit des probabilités des onsets :

$$P_{ioi}(n_1,n_2) = P_{Onset}(n_1) * P_{Onset}(n_2) \tag{8}$$

Les interonsets ne sont pas forcément consécutifs. Un exemple d'histogramme des interonsets est donnée sur la figure 12 (traits pointillés).

5.3 Adaptation de la représentation des onsets à l'audition

Lorsque l'on calcule les interonsets avec la représentation probabiliste donnée au paragraphe 4.2.3, on oublie un facteur d'adaptation de l'oreille. En effet, le tempo du morceau n'est jamais parfaitement régulier. Pourtant, l'oreille corrige cette irrégularité. De plus, la position de l'onset n'est obtenue qu'avec une précision de quelques millisecondes. Nous proposons de modéliser cette correction d'erreur en répartissant la probabilité de l'onset sur une fenêtre temporelle autour de l'onset détecté. Nous avons testé l'estimation du tatum sur une base de données de son percussifs avec une fenêtre rectangulaire, une fenêtre gaussienne et une fenêtre de Hanning. Il s'avère que la fenêtre rectangulaire est celle qui donne les meilleurs résultats. Nous obtenons avec les autres fenêtres des erreurs supérieures à 5 % pour l'estimation du tatum, ce que nous n'avons pas avec notre algorithme (cf 7.3). On convolue donc le vecteur de vraisemblance des onsets détectés par cette fenêtre de tolérance. Un exemple est donné sur la figure 12

5.4 Résolution par transformée de Fourier

Les interonsets théoriques sont par définition des multiples de la micropulsation T_{tatum}. La somme des interonsets est donc un signal non nul seulement aux instants $k * T_{tatum}$. Le signal est en fait un train d'impulsion d'amplitude différentes. L´écart entre les impulsions est T_{tatum}. Sa transformée de Fourier est donc un train d'impulsion $\chi(\omega)$ aux multiples du tatum τ(exprimé en bpm).

L'effet de l'introduction de la fenêtre temporelle (adaptation de la représentation des onsets) est de multiplier la transformée de Fourier par une sinus cardinal (pour une fenêtre carrée). La largeur du lobe principal dépend de la largeur de la fenêtre de départ. Nous utilisons une fenêtre de 60ms. Le premier zéro est atteint pour un tatum de 2000 bpm. Nous cherchons dans la transformation de Fourier des tatums de fréquence maximale de 400 bpm

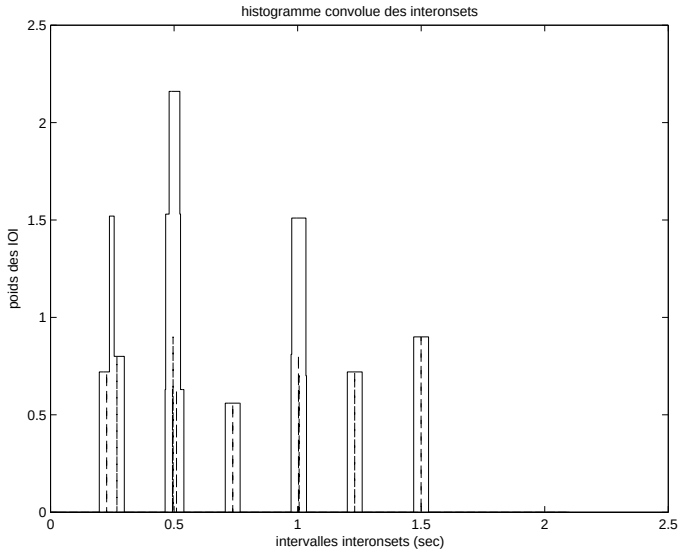

FIG. 12 – Régularisation des interonsets obtenu sur la figure 11. Les interonsets sont les traits en pointillés ; la courbe en trait plein est le lissage de l'histogramme par une fenêtre rectangulaire

donc nous observons le tatum dans le premier lobe. Cette fenêtre a tout de même un effet sur les pics du train d'impulsion $\chi(\omega)$: les pics décroissent. La valeur du tatum est prise comme le premier pic de $\chi(\omega)$, que l'on considère être le pic maximal de la transformée de Fourier. Un exemple est donné sur la figure 13.

Il existe d'autres méthodes comme le TWM utilisé par Fabien Gouyon [GHC02]. Nous nous sommes limités l'utilisation de la transformée de Fourier car elle donnait des résultats satisfaisants(cf 7.3).

5.5 Détection de la phase du tatum

Tous les véritables onsets se trouvent théoriquement sur les microbattements (cf figure 4). Nous choisissons donc de prendre pour microbattement U_m de départ l'onset que l'on a détecté avec la probabilité maximale : $U_m = argmax(P_{onset(n_i)})$. Les autres microbattements se déduisent en ajoutant ou retranchant les multiples de la valeur du tatum de μ : $U_{m\pm k} = \text{reste}(U_m/T_{tatum}) \pm k * T_{tatum}$.

La définition du tatum permet de segmenter la séquence musicale en fenêtre F_m de longueur égale T_tatum centrées sur les micropulsations U_m qui seront des unités d'analyse par la suite. Un exemple est donné sur la figure 15

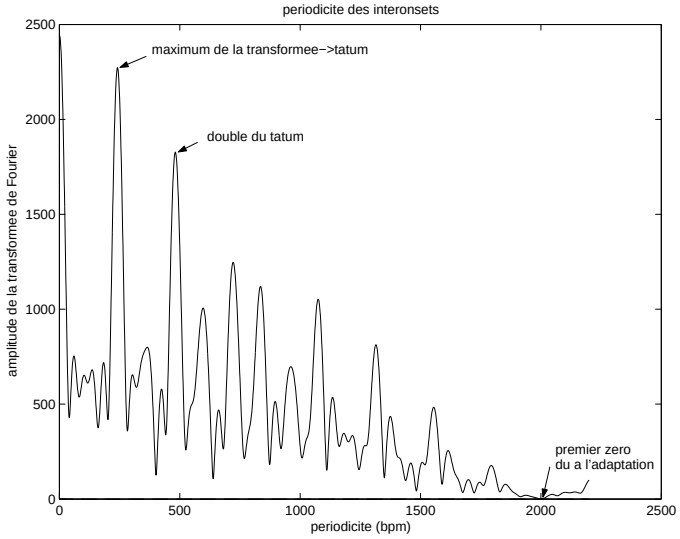

FIG. 13 – Exemple de graphe obtenu en faisant la transformée de Fourier des interonsets convolué. On voit la forme générale en sinus cardinal, les multiples du tatum, et le tatum comme maximum. Ici le tatum est de 240.

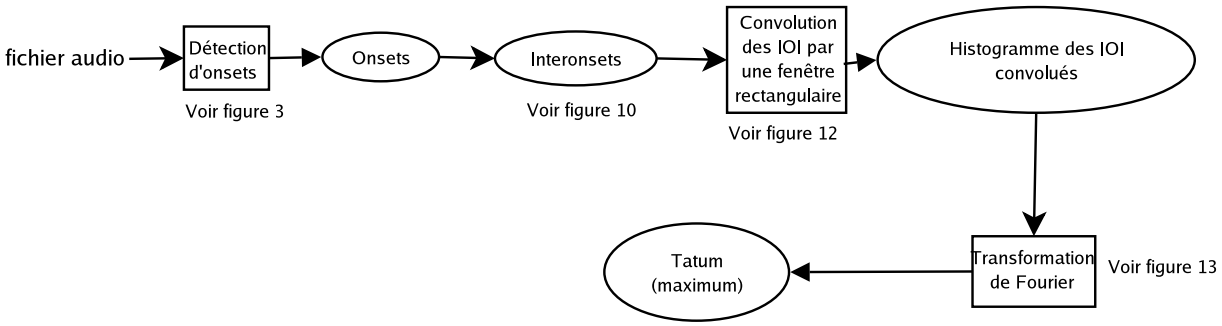

FIG. 14 – Algorithme de détection du tatum

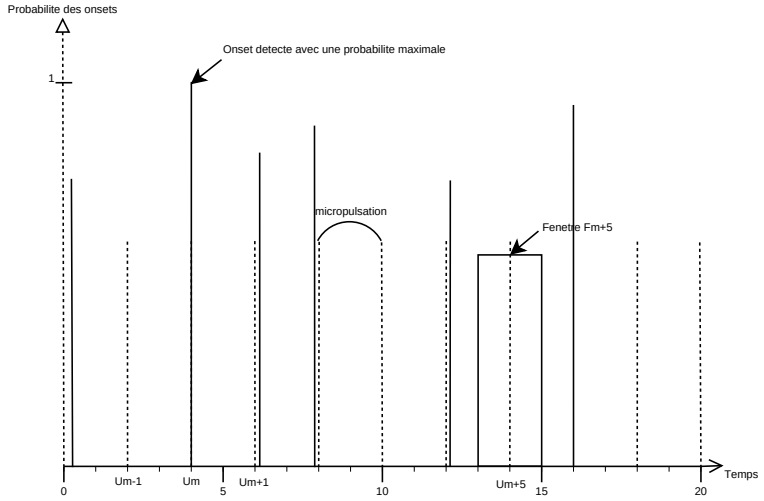

FIG. 15 – Détection du tatum de la phase du tatum : les onsets sont dessinés en noirs. Leur hauteur est égale à leur probabilité. La grille des micropulsations en gris. L'onset à la position U_m détermine la phase du tatum

6 Algorithme de détection de la pulsation

Contrairement à l'algorithme de Scheirer qui estime directement la pulsation à partir du signal, la pulsation est ici déduite du tatum donc des onsets. Nous regardons donc des informations discrètes au sens où l´échelle du tatum est beaucoup plus grande que l'échelle de l´échantillonnage.

6.1 Un nombre limite de pulsations possibles

La micropulsation étant le plus grand commun diviseur des interonsets observé dans le morceau, la pulsation est un multiple de la micropulsation. La micropulsation est donc une information intéressante pour retrouver la pulsation. Il reste à savoir quel multiple choisir.

6.2 Définition de la pulsation/tactus à partir du tatum

Lorsque l'on écoute une séquence, on pourrait seulement la caractériser par le tatum. Or, il s'avère que la pulsation et la mesure sont beaucoup plus faciles pour un humain à repérer. On peut se demander alors quels sont les choix faits par notre cerveau pour s'accorder sur la pulsation. [Meu04] met en avant l'accentuation périodique de certaines pulsations ou micro-pulsations. Il s'agit pour lui afin de retrouver le tempo et la mesure de séparer les temps forts des temps faibles.

Nous avons fait des tests en prenant comme temps fort les micropulsa-

tions U_m où la probabilité d'un onset est grande, et comme temps faible les micropulsations U'_m où aucun onset n'est détecté. Nous ne sommes pas arrivés à dégager une caractérisation de la pulsation satisfaisante. Séparer les temps forts des temps faibles semble n'être qu'une composante de la régularité (ou notre définition de la probabilité d'un onset n'a pas suffisamment de sens). Il peut y avoir une régularité de notes (prélude de Bach), de timbre (boucle de batterie), de rythme (comme cas extrême de la séparation temps fort-temps faible). Il semble alors difficile de prendre la méthode de Benoît Meudic telle quelle. Nous utilisons une périodicité selon plusieurs dimensions appropriées : le signal donne beaucoup plus d'informations qu'une simple probabilité d'onset pour chaque U_m. Nous essayons dans la suite de combiner ces informations avec notre représentation probabiliste.

6.2.1 Amélioration de la détection des onsets

Les interonsets ne sont pas aussi réguliers que la micropulsation théorique. Il peut y avoir un décalage entre un onset et sa micropulsation correspondante (par exemple lors d'un ralenti ou d'un accelerando). Pour cela, on cherche autour des micropulsations calculées par l'algorithme de recherche de tatum décrit dans 5.5, l'événement (ou onset) le plus probable. La probabilité de l'onset est pondérée par un coefficient de déviation qui est d'autant plus grand que l'onset est proche de l'onset théorique :

$$Onset_m \quad = \quad max(P_{onset}(n) * f_{U_m}(n), n \in U_m) \tag{9}$$

où $f_{U_m}(n) = \exp(-(\frac{n}{U_m/4})^2)$ a été pris comme une gaussienne d´écart-type égale à un quart du tatum.

Cette méthode utilisée pour des trames de 3 sec ne suffit dans le cas dans le cas d'une variation de tempo à l'intérieur de la trame : la grille fixe du tatum est loin de la réalité. Il vaut mieux alors utiliser une méthode propagative [Pee01] : On cherche onset après onset l'onset le plus probable. Lorsque l'on a calculé cet onset on cherche le suivant dans une fenêtre centrée un microbattement plus tard. Le problème de cette méthode est l´éloignement du tatum constant. Sur nos exemples, la deuxième méthode reste plus efficace. Un exemple d'analyse est donné sur la figure 16.

Ainsi, pour chaque micropulsation est trouvé l'endroit le plus probable où se situe un onset. Le calcul des valeurs de paramètres aux endroits pertinents que sont les début des onsets.

6.2.2 Étude de paramètres propices à la détection du beat

L'étude porte essentiellement sur des sons percussifs. Nous nous concentrerons sur des paramètres caractéristiques de ces sons. Nous calculons ses paramètres sur des fenêtres de Hanning F_m de 10 ms (longueur moyenne

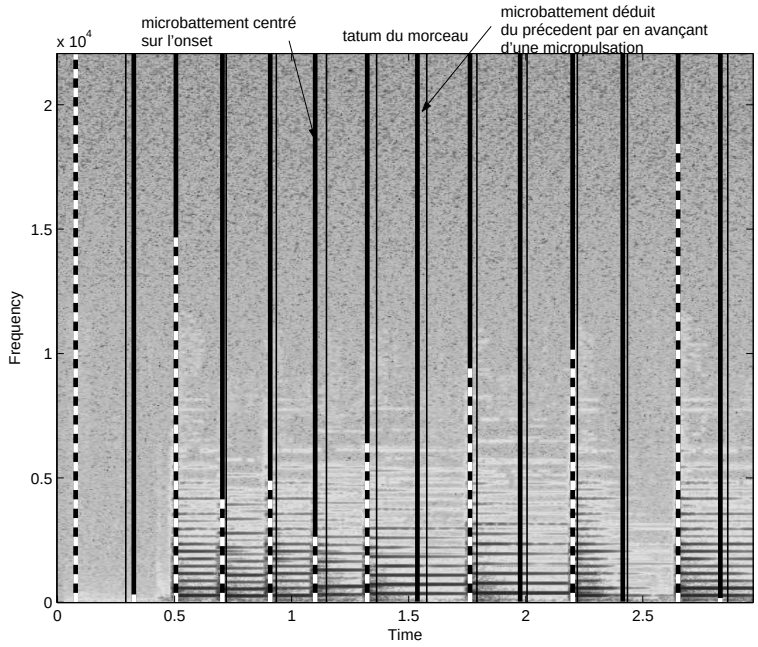

FIG. 16 – La grille théorique du tatum est marquée en traits fins noir, les onsets (sélectionnés) en pointillés blancs. Lorsqu'aucun onset significatif (huitième micropulsation) n'est détecté la micropulsation est placée une micropulsation plus tard par rapport au microbattement précédent. En revanche, lorsqu'il y a un décalage, la micropulsation est placée au temps de l'onset (sixième micropulsation)

approximative d'un son percussif). Le début d'une fenêtre F_m est la micropulsation U_m (cf 6.2.1). Les trois paramètres que nous avons essayés sont le centroïde spectral, l´étalement spectral et l´énergie. L'énergie est la somme du carré du signal.

Le centroïde spectral cs est le barycentre spectral de l´énergie [Pee04] :

$$cs = \frac{\sum_{b=1}^{nbfiltre} b * energie(b)}{\sum_{b=1}^{nbfiltre} energie(b)} \tag{10}$$

où nbfiltre est le nombre de filtre de ERB et energie(b) l´énergie du signal filtré par le bième filtre de ERB.

L´étalement spectral es est la variance de l´énergie :

$$es = \sqrt{\frac{\sum_{b=1}^{nbfiltre} (b - cs) * energie(b)}{\sum_{b=1}^{nbfiltre} energie(b)}} \tag{11}$$

Ils permettent en particulier de séparer la grosse caisse (centroïde bas, étalement faible) de la caisse claire et de la cymbale (centroïde haut, grand étalement). L´énergie est elle différente entre les segments où un onset est détecté et celles où il n'y en a pas. L'autre avantage du centroïde et de l´étalement spectral est qu'ils sont bornés dans la plage de fréquences entre 0 et la demi-fréquence d'échantillonnage donc plus facile à représenter.

6.3 Étude de l'autocorrélation

La mesure ou le battement sont supposés produire une périodicité de la fonction d'observation de chaque paramètre (énergie, centroïde spectral, étalement spectral). Nous cherchons donc cette périodicité. Puisque la mesure est un multiple entier du tatum (que nous connaissons), il suffit de mesurer la corrélation de la fonction d'observation de chaque paramètres aux multiples entiers du tatum. Nous considérons que tous les paramètres ont la même importance. Nous normalisons donc l'autocorrelation pour chaque paramètre et nous observons la somme des trois autocorrélations. La figure 18 montre un exemple de calcul de cette fonction. Nous ne sommes pas parvenus à développer un algorithme fiable d'extraction des pics de la fonction d'autocorrélation, afin de retrouver le beat. Nous obtenons seulement des pics à des périodicités significatives.

7 Expériences et résultats

7.1 Base de données de test

Base de données de sons percussifs ou assimilés La base de données est constituée de 65 extraits sonores d'une durée de quelques secondes

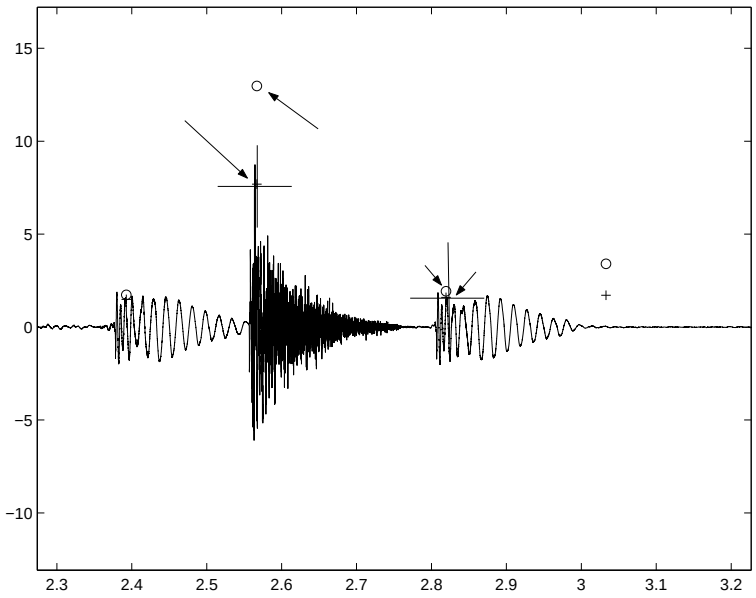

FIG. 17 – Comparaison entre les valeurs du centroïde (croix) et de l´étalement spectral (rond) pour une cymbale à gauche et une grosse caisse à droite.

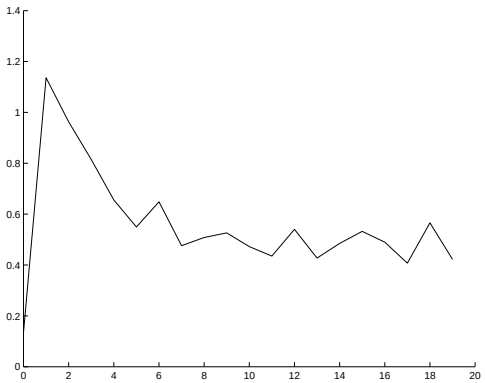

FIG. 18 – periodicité obtenue pour 10 secondes de "waterloo" de ABBA. Le morceau est un 6/8 et la micropulsation est la croche. Le graphique obtenu des pics forts d'autocorrélation pour les multiples de la mesure et des pics faibles pour les multiples impairs (9,15) de la pulsation

(de 3 à 10 sec). Les sons sont disponibles à l'adresse :
http ://www.tplm.com/ samples /boucles/groove/home.htm Ce sont des sons de percussions réelles ou électroniques. Les onsets ont été indexés manuellement par l'observation des spectrogrammes et l'écoute des extraits sonores. L'ensemble des extraits contient 912 onsets. Le tactus était déjà indexé. Le tatum a été indexé à partir du tactus en faisant le rapport entier entre l'intervalle minimal entre les onsets et la pulsation. $\tau = tactus * arrondi(\frac{IOI_{tatum}}{IOI_{tactus}})$.

Extraits sonores de musique pop La base de données est constituée d'une dizaine d'extraits de morceaux de musique pop et d'un morceau de jazz. Ce sont des extraits variant entre 10 et 20 secondes de

1. Message in the bottle de Police
2. Paranoïd de Radiohead
3. You are the sunshine of my life de Stevie Wonder
4. Waterloo de ABBA
5. Dancing Queen de ABBA
6. Hotel California des Eagles
7. Angie des Rolling Stones
8. Take Five de Dave Brubek
9. extraits du CD Standards of excellence 2

Ces extraits contiennent aussi un rythmique basée sur des percussions réelles ou virtuelles.

7.2 Evaluation de l'algorithme de détection des onsets

Nous avons évalué l'algorithme de détection des onsets de sons percussifs. Notre critère d'évaluation, donné dans [Kla99], est le suivant :
efficacite = $\frac{\alpha-\beta-\gamma}{\alpha}$ où α est le nombre théorique d'onsets, β le nombre de fausses détections et γ le nombre d'onsets non détectés.

L'efficacité est de 1 s'il n'y a pas d'erreur et diminue d'autant plus que des onsets ne sont pas détectés où que des onsets sont faussement détectés. La figure 19 représente la mesure de cette efficacité en fonction du seuil de l'indice de confiance. On voit que l'on obtient 87% de bonnes détections pour un seuil fixé à 10%. Ce résultat est comparable à celui observé dans [Kla99] Pour ce seuil, le nombre d'onsets non détectés est de 36 et le nombre d'onsets faussement détectés est de 60 (le nombre d'onsets était de 912).

7.3 Evaluation de l'algorithme du tatum

Sur la même base de tests que pour 7.2, nous avons estimé la fiabilité du tatum avec une tolérance de 1 % (cf [GHC02]). La tolérance est définie par : $tolerance = \frac{|tatumreel-tatumtrouve|}{tatumreel}$. Sur les 63 extraits sonores,

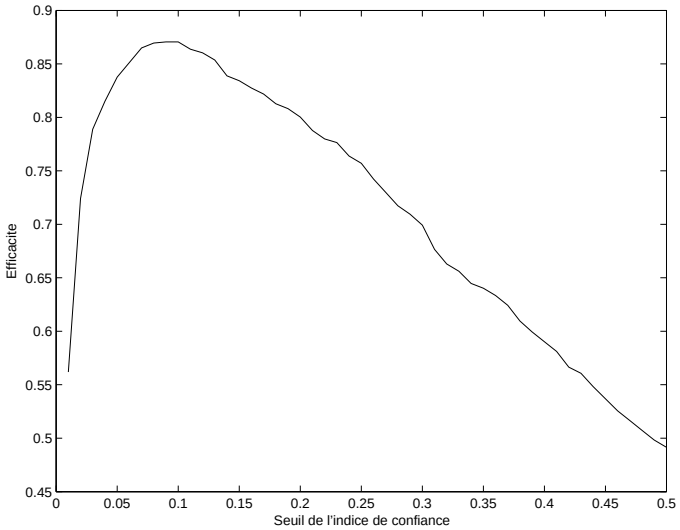

FIG. 19 – Efficacité en fonction du seuil de l'indice de confiance

1. 54 tatum sont correctes (avec une erreur inférieure à 1%),
2. 3 tatums sont évalués comme la moitié des tatums réels, un comme le double.
3. Les 5 derniers fichiers sont faux avec une erreur variant entre 1% et 2%. Cela reste peu, et en utilisant l'algorithme propagatif (pour trouver les microbattements), nous parvenons à réajuster le tatum. L'erreur est audible à la fin de l'extrait si on ne la corrige pas.

L'algorithme est sensible à certaines fausses détections et non-détections, par exemple si la note la plus courte n'est pas détectée. Ce sont les causes des mauvaises évaluations du tatum. Les fausses détections rajoute un niveau à la métrique. Les non-détections en enlèvent un.

N.B. : Le travail sur cette base de données a soulevé un autre problème non résolu, mais qui n'apparaît pas dans les résultats : Le décalage dû au groove peut fausser l'estimation du tatum (voir figure 20)

Nous avons faits des tests sur la base de données de musique pop. Malgré un algorithme de détection d'onsets moins fiable, nous obtenons une estimation du tatum correcte pur la plupart des fichiers. Les seules erreurs sont dues à des fausses détections qui multiplient la micropulsation par 2. D'autre part, les tests de suivi de tatum sont aussi encourageants. L'algorithme est capable de suivre un ralenti ou un accelerando. Nous ferons écouter des exemples lors de la soutenance.

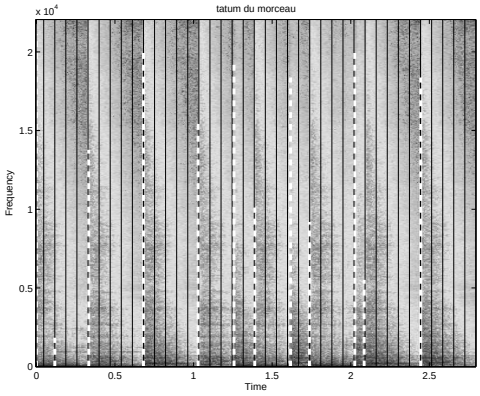

FIG. 20 – Exemple de fausse appréciation du tatum : le battement est divisé en 5 et non en 2. Certains onsets sont décalés et le phénomène est accentué par 2 fausses détections.

7.4 Présentation de résultats relatifs à la détection du tempo et de la mesure

L'algorithme relatif à la description haut niveau de la métrique n'est pas terminé. Nous avons tout de même pu tester l'usage de l'autocorrélation sur la bases de données de musique pop. Les maxima de la fonction d'autocorrélation coïncident avec certains niveaux de la métrique pour tous les fichiers. Il s'agit de la hiérachie des fréquences associées à l'énergie lorsque toute la grille des microbattements n'est pas remplies ou des périodicités de certains instruments percussifs très remarquables dans le morceau. Des exemples sont données à la fin du rapport. En revanche, nous n'avons pas trouvé d'algorithme général pertinent pour choisir le pic d'autocorrélation correspondant à la pulsation ou à la mesure. Notre caractérisation n'est surement pas assez fine pour pouvoir aller plus loin.

8 Conclusion et perspectives

L'algorithme développé au cours de ce stage permet la conception d'une grille de microbattements sur laquelle sont situés tous les événements rythmiques. Cette grille tient compte des variations de tempo. L'étude de descripteurs simples sur cette grille permet sur des enregistrements de musique qui contiennent des percussions de retrouver une partie de la hiérarchie des périodicité rythmiques. La première amélioration à apporter est de terminer l'algorithme de détection du tempo . Pour cela, il reste une étude à faire sur les paramètres perceptivement pertinents pour la recherche de périodicité pour tous les types de fichier. Cela sous-entend aussi le développement d'une

fonction de détection des onsets adaptée. Enfin, l'histogramme des intervalles inter-onsets est un sujet d´étude intéressant pour la recherche des périodicités associées au rythme. Nous avons pu grâce à une représentation probabiliste des onsets introduire une pondération des éléments de cet histogramme. Cependant, nous pensons qu'en introduisant des critères de similarité entre événements (événements de timbre proche) dans la définition de l'histogramme, nous pourrions obtenir un IOI-gramme représentatif des pulsations présentes dans l'extrait sonore à chaque instant.

Abréviation	nom complet	unité
IOI	intervalle inter-onset	sec
n	indice temporel	
m	indice relatif aux microbattements	
b	indice relatif aux bandes de fréquences	
i	indice relatif aux onsets	
$P_{Onset}(n)$	Probabilité d'onset à l'échantillon n	sans
$f_{Onset}(n)$	Fonction de détection d'onset à l'échantillon n	
$F_{Onset}(n)$	Fonction d'apparition d'un onset à l'échantillon n	
T_{tatum}	micropulsation	sec
	pulsation	sec
τ	tatum	bpm(60*Hz)
	tempo	bpm(60*Hz)

TAB. 2 – Tableau des abréviations et des unités

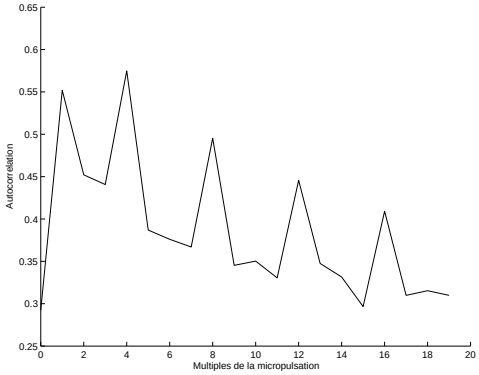

FIG. 21 – Périodicité obtenue pour 10 secondes de "You are the sunshine of my life" de Stevie Wonder. Le morceau est un 4/4 et la micropulsation est la double-croche. Le graphique obtenu des pics forts d'autocorrélation pour les multiples de la pulsation (4,8,12,16). En revanche, on ne voit pas un accentuation de la mesure (16) par rapport à la pulsation

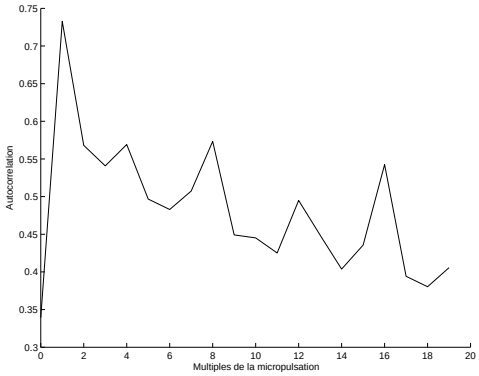

FIG. 22 – Périodicité des paramètres aux multiples de la micropulsation obtenue pour 10 secondes de "Message in the Bottle" de Police. Le morceau est un 4/4 et la micropulsation est la double-croche. Le graphique obtenu des pics forts d'autocorrélation pour les multiples de la mesure (8,16) et des pics faibles pour les multiples pairs (4,12) de la pulsation

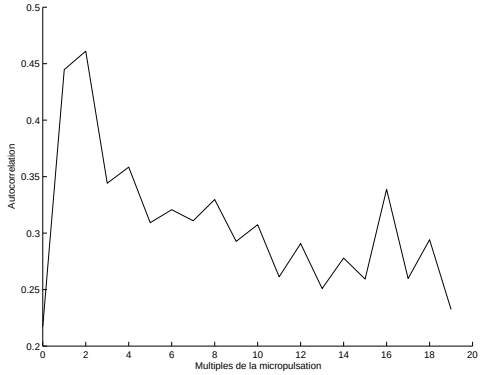

FIG. 23 – Périodicité des paramètres aux multiples de la micropulsation obtenue pour 10 secondes de "Paranoïd" de Radiohead. Le morceau est un 4/4 et la micropulsation est la double-croche. On obtient des pics forts d'autocorrélation pour le multiple de la mesure (16) et un pic assez fort pour la longueur correspond à la blanche (8), des pics plus faibles pour la pulsation (4, 12) et des pics encore plus faibles pour les croches

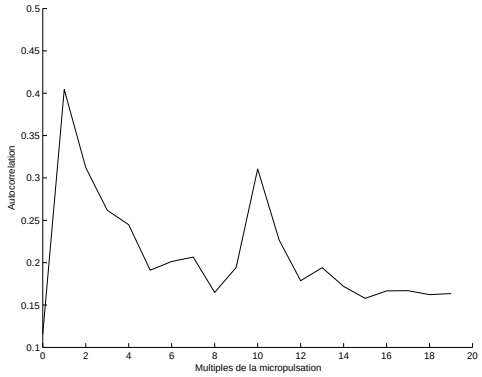

FIG. 24 – Périodicité des paramètres étudiés aux multiples de la micropulsation obtenue pour 10 secondes de "Take Five" de Dave Brubek (sans le saxophone solo). Le morceau est un 5/4 et la micropulsation est la croche. On obtient un pics très fort d'autocorrélation pour la mesure(10).

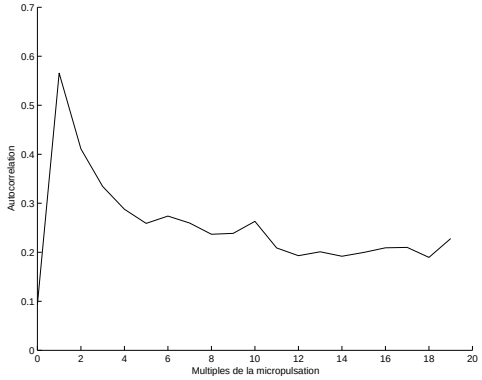

FIG. 25 – Périodicité des paramètres étudiés aux multiples de la micropulsation obtenue pour 10 secondes de "Take Five" de Dave Brubek (avec le saxophone solo). Le morceau est toujours un 5/4 et la micropulsation est toujours la croche. Le pic d'autocorrélation pour la mesure est toujours présent même s'il est moins fort

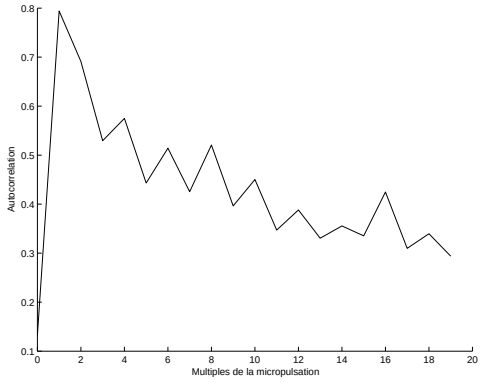

FIG. 26 – Somme des autocorrélations de chaque paramètre étudié aux multiples de la micropulsation obtenue pour 10 secondes de "Dancing queen" de ABBA. Le morceau est un 4/4 et la micropulsation est la double-croche. On obtient des pics forts d'autocorrélation pour le multiple de la mesure (16) et un pic assez fort pour la longueur correspond à la blanche (8), des pics plus faibles pour la pulsation (4,12) et des pics encore plus faibles pour les croches

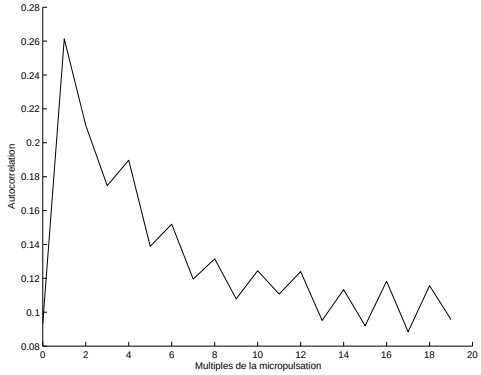

FIG. 27 – Somme des autocorrélations de chaque paramètre étudié obtenue aux multiples de la micropulsation pour 10 secondes de "Hotel California" des Eagles. Le morceau est un 4/4 et la micropulsation est la double-croche. On obtient des pics forts d'autocorrélation pour les multiples de la croche, en revanche on ne peut deviner des structures plus grandes.

Références

[ABDR03] M. Alonso, R. Badeau, B. David, and G. Richard. Musical tempo estimation using noise subspace projections. IEEE, 2003. Workshop on Applications of Signal Processing to Audio and Acoustics.

[Bil93] J.A. Bilmes. Timing is of essence. Master's thesis, Massachusetts Institute of technology, 1993.

[BN93] M. Basseville and I. Nikiforov. Detection of abrupt changes : theory and application, 1993.

[Chi95] M Chion. *Guide des objets sonores*. INA,GRM, novembre 1995.

[DS04] J.P. Bello L.Daudet S. Abdallah C. Duxbury M. Davies and M.B. Sandler. A tutorial on onset detection in music signals. IEEE, 2004.

[GHC02] F. Gouyon, P. Herrera, and P. Cano. Pulse dependent analysis of percussive music. AES, 2002. Conference on Virtual,Synthetic and Entertainment Audio.

[HDG03] P. Herrera, A. Dehamel, and F. Gouyon. Automatic labeling of unpitched percussion sounds, 2003.

[J.B93] J.Maher J.Beauchamp. Fundamental frequency estimation of musical signals using a two-way mismatch procedure. *Journal of the Acoustical Society of America*, 1993.

[JR01] F. Jaillet and X. Rodet. Detection and modeling of fast attack transients, 2001.

[Kla99] A. Klapuri. Sound onset detection by applying psychoacoustic knowledge. IEEE, 1999. Conf. Acoustics,Speech and signal processing.

[KP02] A. Klapuri and J. Paulus. Measuring the similarity of rhythmic patterns. IRCAM, 2002.

[Lad89] C.K. Ladzekpo. Foundation course in african dance drumming. U.C. Berkeley, Department of Music, 1989, 1989.

[Meu04] Benoit Meudic. *Détermination automatique de la pulsation, de la métrique et des motifs musicaux dans des interprétations à tempo variable d'oeuvres polyphoniques*. PhD thesis, université Paris VI, 2004.

[MG83] B Moore and B Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *JASA*, 74(3) :750–753, September 1983.

[P F74] P Fraisse. *Psychologie du rythme*. PUF, 1974.

[Pee01] Geoffroy Peeters. *Modèles et modélisation du signal sonore adaptés à ses caractéristiques locales*. PhD thesis, université Paris VI, 2001.

[Pee04] Geoffroy Peeters. A large set of audio features for sound description(similarity and classification) in the cuidado audio project. 2004.

[Sch98] E. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1) :588–601, 1998.

[Sep01] J. Seppänen. Computational problems of musical meter recognition. Master's thesis, Tampere University of Technology, 2001.

[URCH04] C. Uhle, J. Rohden, M. Cremer, and J. Herre. Low complexity musical meter estimation from polyphonic music. London,United Kingdom, June 2004. AES.