

Université Paris VI - Pierre et Marie Curie
École Doctorale EDITE

Thèse en vue de l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité
Acoustique, Traitement du Signal, Informatique, Appliqués à la Musique

Suivi de Tempo Appliqué aux Musiques Improvisées
à la recherche du temps perdu ...

Gilbert NOUNO

Cette thèse a été soutenue le Jeudi 20 Novembre 2008
devant le jury composé de

Philippe CODOGNET	Directeur
Marc CHEMILLIER	Rapporteur
David WESSEL	Rapporteur
Carlos AGON	Examineur
Gérard ASSAYAG	Examineur
Miller PUCKETTE	Examineur

Thèse préparée à
l'IRCAM – CNRS UMR 9912
Equipe Représentations Musicales
1, place Igor Stravinsky, 75004 Paris.
<http://www.ircam.fr/>

*À mes parents, pour la passion de la musique qu'ils m'ont donnée.
À Jean-Claude Risset, avec qui tout commençait un jour de juin de l'année 1993.*

Remerciements

Cette thèse est le fruit de plusieurs années de travail à l'Ircam, partagées entre les départements de la production musicale et de la recherche. Ce travail n'aurait sans doute jamais vu le jour sans l'inspiration et le soutien des nombreuses personnes que je tiens à remercier chaleureusement.

Merci à mes parents pour tout ce qu'ils représentent dans mon parcours.

Merci à Gérard Assayag de m'avoir accueilli au sein de l'équipe *Représentations Musicales*, à Raphaël et Carlos Agon pour leur soutien constant durant de longs week-ends, à Moreno Andreatta, à Mondher Ayari, à Jean Bresson pour ses conseils avisés, à Arshia Cont pour son partage aussi musical que scientifique, à Philippe Codognet pour sa patience, à Marc Chemillier pour l'appréciation de ma recherche, à Olivier Warusfel pour sa complicité musicale et à Jean Laroche pour ses recommandations éclairées.

Merci à Alain Jacquinot pour le temps qu'il a bien voulu me concéder lors de la réalisation de cette thèse, à Hugues Vinet et Bernard Stiegler de m'avoir accordé leur confiance.

Merci à David Wessel et à Miller Puckette pour leurs présences modèles.

Merci à Bruno Simeone pour son aide infiniment précieuse lors de la formalisation de ce travail.

Merci aux compositeurs et musiciens qui ont été et restent la plus grande source d'inspiration, Hacène Larbi, Malik Mezzadri, Patrick Moutal, Jonathan Harvey, Philippe Schoeller, Michaël Levinas, Marco Stroppa, Octurn, Aka Moon, Riccardo Del Fra, Ravi Coltrane, Steve Coleman et Pierre Boulez.

Enfin, merci à Chiara pour sa présence à mes côtés et tout ce qu'elle est pour moi.

Gilbert Nouno, Paris, Novembre 2008

“On a pu, pendant des millénaires, faire usage du son pour chanter ou pour jouer et danser. Ce fut la première musique qui se transmettait oralement, sans plus.

Mais un jour – six siècles av. J.-C – quelqu’un s’inquiète de rendre transmissible pour toujours, l’une de ces musiques autrement que de bouche à oreille, donc de l’écrire. Il n’existait ni méthode ni instrument pour le faire. Il s’agissait de fixer ce son en des points déterminés, rompant ainsi sa parfaite continuité. Il fallait le représenter par des éléments saisissables, par conséquent découper *le continu* selon une certaine convention et en faire *du gradué*.

Comment sectionner la continuité du phénomène sonore ? Comment découper ce son selon une règle admissible par tous, mais surtout efficace, c’est-à-dire capable de souplesse, de diversité, de nuances et de richesse et pourtant simple et maniable et accessible ?

Pythagore résolut la question, prenant deux points d’appui capables de rallier la sécurité et la diversité : d’une part, l’oreille humaine – l’audibilité humaine. D’autre part, les nombres, c’est-à-dire la mathématique (ses combinaisons) qui est elle-même fille de l’Univers.

Ainsi fut créée la première écriture musicale capable d’enfermer des compositions sonores et de les transmettre à travers temps et espace. Sait-on qu’en ce qui concerne les choses visuelles, les *longueurs*, nos civilisations n’ont pas encore franchi l’étape accomplie par la musique ? Tout ce qui est bâti, construit, distribué en longueurs, largeurs ou volumes, n’a pas bénéficié d’une mesure équivalente à celle dont jouit la musique, – outil de travail au service de la pensée musicale.”

Le Modulor
LE CORBUSIER

Table des matières

Remerciements	v
Liste des figures	xiv
Liste des algorithmes	xv
Liste des tableaux	xvii
Abstract and introduction	xix
Résumé	xxv
1 Introduction	1
1.1 Interaction temporelle musique et informatique	2
1.2 Ethnologie numérique	3
1.3 Cheminements	5
1.4 Éléments d'analyse musicale, temps et culture	6
1.5 Tempo, pulsation, tactus et tatum	7
1.6 Continuité, discontinuité : une arche du temps musical	8
1.7 Anticipation musicale	9
1.8 Plan de lecture de la thèse	10
2 Suivi de tempo : fondements et état de l'art	13
2.1 Applications musicales	14
2.2 Quelques définitions	14
2.3 Identification des problématiques	15
2.4 Prétraitement du signal musical	16
2.5 Variables observables et non observables, hypothèses	18
2.6 Perception du temps musical : études psychocognitives	20
2.7 Etat de l'art	21

3	Grille optimale et estimation de tempo	25
3.1	Grille de quantification	26
3.1.1	<i>Tactus</i> ou <i>pas</i> de la grille de quantification	27
3.1.2	Résolution géométrique	28
3.1.3	Remarques	35
3.2	Meilleure phase, grille de quantification optimale	37
3.2.1	Reformulation du problème	37
3.2.2	Périodicité et phase optimales	38
3.2.3	Algorithme de minimisation de la phase de la grille	42
3.2.4	Domaine de définition de la fonction à optimiser	43
3.2.5	Coïncidence optimale	45
3.2.6	Séquence temporelle et repliement	45
3.3	Résolution bidimensionnelle	46
3.3.1	Paramétrisation de la grille optimale	47
3.3.2	Algorithme d'optimisation bidimensionnelle	48
3.3.3	Recherche du meilleur nombre de points de grille : indice de coïncidence	55
3.3.4	Réduction de l'algorithme d'optimisation	60
3.3.5	La grille optimale contient deux points d'entrée	62
4	Accélération de tempo et grille optimale	69
4.1	Tempo uniformément et continûment accéléré	70
4.1.1	Notation et interprétation, transcription et performance : la ques- tion du temps en musique	70
4.1.2	Modélisation du tempo par une pulsation accélérée uniformément	73
4.1.3	Accélération directe uniforme	76
4.1.4	Accélération réciproque uniforme	78
4.2	Coïncidence optimale, séquence à tempo non constant	83
4.2.1	Modèle de grille optimale uniformément accélérée	84
4.2.2	Propriétés de la fonction $\psi(\theta)$	85
4.2.3	Propriétés de la fonction $\Phi(s, \Delta, \theta)$	90
4.3	Algorithme de minimisation de $\Phi(s, \Delta, \theta)$	91
4.3.1	un schéma itératif alterné	91
4.3.2	Première étape de minimisation	92
4.3.3	Seconde étape de minimisation	93
4.3.4	Algorithme généralisé pour minimiser $\Phi(s, \Delta, \theta)$	95

4.4	Utilisation musicale de la grille optimale	97
4.4.1	Transcription rythmique et modèle probabiliste	98
5	Suivi de tempo	101
5.1	Inférence et modèle probabiliste	102
5.1.1	Modèle d'espace d'états	103
5.1.2	Modèle probabiliste de tempo	104
5.1.3	<i>HMM</i> et décodage Viterbi	107
5.2	Synchronisation : un modèle d'oscillateur non linéaire	110
5.3	Synchronisation et prédiction de la pulsation	117
5.4	Implémentation temps réel	120
5.5	Expérimentations et discussion	123
6	Conclusions et extensions futures	125
6.1	Conclusions des recherches en cours	125
6.2	Recherches futures et extensions	128
A	Interviews	131
B	Elements de Probabilité	147
B.0.1	Elements de probabilité	147
C	Raisonnements probabilistes	157
C.0.2	Raisonnements probabilistiques	157
D	Publication d'article	167
	Bibliographie	185

Table des figures

2.1	Schéma général du suivi de tempo, de la source au suivi	17
3.1	exemple de grille avec $k = 3$ et $p = 5$	28
3.2	$f_k(\Delta)$	30
3.3	$g(\Delta)$	31
3.4	zoom de $g(\Delta)$	31
3.5	$g(\Delta)$ avec $p = 4$	32
3.6	Points singuliers pour une séquence donnée	36
3.7	Grilles de 100 et 101 ms pour la séquence (100, 203, 406, 702, 989, 1310)	37
3.8	$m_k = 3$ et $M_k = 5$ pour le calcul de $\varphi(s)$	40
3.9	$\lambda_{hk}(s)$	40
3.10	$\mu_k(s)$	41
3.11	Domaine de définition pour minimiser $\varphi(s, \Delta)$	44
3.12	Approximation de f sur $\mathcal{D}(s, \Delta)$	44
3.13	p -grille paramétrée sur q et l	47
3.14	Meilleur nombre de points de grille	55
3.15	fonction <i>seawave</i>	57
3.16	points anguleux et intervalles constants de ρ	59
4.1	Séquence rythmique accélérée uniformément	79
4.2	accélération directe uniforme d'une séquence de 20% , $\alpha = 1.2$	79
4.3	courbes de tempo	82
4.4	accélération réciproque $\psi(\theta, \tau)$	86
4.5	Accélération réciproque $\psi(\theta, \tau)$ pour τ fixé	87
4.6	algorithme itératif Alg.4.5 sur une séquence supposée accélérée, on remarque que dès la première itération de θ la suite de point est proche de la suite limite de convergence	97
4.7	Grille Optimale pour les points de la séquence supposée non accélérée	97

4.8	Grille Optimale, séquence supposée accélérée de $\alpha = 1.1$	98
4.9	Grille Optimale, séquence supposée accélérée de $\alpha = 1.2$	98
5.1	modèle d'état <i>HMM</i>	104
5.2	exemple d'observations et transitions d'états associées	109
5.3	inférences du tempo	111
5.4	Pulsation modélisant un champs de receptivité temporel	114
5.5	Distribution de von Mises, coordonnées cartésiennes et polaires	114
5.6	Distribution de von Mises	116
5.7	Adaptation de phase de la pulsation attentionnelle	121
5.8	Adaptation de phase de la pulsation prédite	122
5.9	antepulse dans l'environnement temps réel <i>MaxMsp</i>	123

List of Algorithms

3.1	Minimisation de f sur I	34
3.2	minimum de $\varphi(s)$ pour un Δ donné	43
3.3	Meilleure grille d'approximation bidimensionnelle	49
3.4	Meilleur pas	54
3.5	meilleur pas avec erreur normalisée	60
3.6	Meilleure grille d'approximation bidimensionnelle	61
4.1	accélération uniforme directe d'une séquence temporelle	78
4.2	accélération réciproque uniforme	83
4.3	descente alternative pour la minimisation de $\Phi(s, \Delta, \theta)$	92
4.4	algorithme de Newton pour trouver la racine θ_{hk} de (ξ_{hk}) (4.2.39) . . .	93
4.5	minimisation de $\Phi(s, \Delta, \theta)$	96
5.1	modèle d'oscillateur non linéaire avec <i>focus attentionnel</i>	118

Liste des tableaux

2.1	<i>Beat Tracking</i> : algorithmes principaux	23
-----	---	----

Abstract and introduction

On Improvised Music

This thesis, entitled *Tempo follower applied to improvised music*, is focused on pulsated musical time. By this, we mean a musical time that inductively creates a rhythmic feeling and carries a pulse. The reference to *improvised music* might appear too vague or too selective. Improvisation conjures spontaneity, a music full of surprise for the listener as much as for the performer. Improvisation expects on the part of the listener a certain appreciation of its cultural context. For example, some music might appear improvised for some people but not for other people who understand or derive its codes from listening. But primarily *improvised music* is an exposed music, whether or not it is composed. It is its characteristically spontaneous creation or generation rather than its categorizability that holds our attention. From a computer point of view, the process of “vision” or “listening” are about capturing and processing information. We set our research in the following basic and purposefully simple framework : the computer “listens” to temporal musical information for which no *a priori* clues are given. Hence, for the computer, music appears as an improvisation, a generated information signal which can be observed with an algorithmic process in time.

Musical time and tempo tracking

The term tempo as we are using it refers to a particular conception of musical time. Through the abstract term *musical time*, we are interested in the drifting of some performing time in respect to some symbolic time ; for instance that could be embedded in a score. Symbolic time may or may not exist as a real score or in the mind of the musician, but eventually it always generates the performance. The purpose of our thesis is the synchronization of a dual temporality, a dual musical time : performed

and scored. We propose an original tempo estimation paradigm based on an *optimal grid* search approach (3) : the regular time-spanned grid of beats or pulsation that best coincides with the observed incoming musical events. We achieve a *tempo follower* – also known as *tempo tracker* – within a framework of dynamic inferences (5) combined with an *attending focus* function (5.2) derived from a paradigm of psycho-cognitive research. We design a *tempo follower* algorithm within a real time musical environment. Our intention is to maintain within this framework a relevant human-computer relation so as to feed in both ways the complexity of forms and thoughts of music at its proper creation time.

Music and Computer Science

Music, science, and technology have, perhaps since their conception, been inter-related [Luening, 1968]. In the last century, within the context of music and computer science, new interactions have emerged, each one opening some unexplored or singular regions, on as much of a musical level as a scientific and technological one. This meeting, under the name *Computer Music*, has fed artistic and scientific research for more than half of a century. It has joined spaces that might have appeared unrelated at first. *Computer Music* is not a musical style but rather the gathering of two living heritages, artistic and scientific [Puckette, 2002]. One must decide by himself and through his own experience if the result will be an artistic science, a scientific art or merely a new cultural field – and to a certain extent social, economic and politic – which surpasses the labels people wish to attach [Chadabe, 2000].

Time, music and computers

Among the fields of interaction which lie in the domain of *Computer Music*, our research focuses on the temporal domain. We consider questions of synchronization in music, particularly synchronization in real time between performing musicians and computers. We do not aim to create a *tempo tracker* for mere technological purposes. A *tempo tracker* can be reductive as it might point out at different meanings in respect to what it does : it could be global tempo estimation in a music piece, it could provide automatic synchronization of music sequences with constant tempo or it could look at tempo information to categorize a musical database, etc. Many different behaviors could be given the label *tempo tracking* ; it is among these behaviors that we must de-

cide our position. We define our research in a causal and real time framework. We are interested in elaborating upon and using intelligent rhythmic agents in the dynamic development of a complex musical environment.

Towards an ideal of a computer-musician synchronization

Why synchronization in music? Even before considering questions of esthetics – in respect to playing together at the same time – synchronization originates from a desire of communication, a desire to share something, an emotion, within the dynamic of time. Synchronization underlies mutual listening and makes musicians’ participation active between them : music *ensembles* – meaning together – is a word of evidence for this.

From avant-garde jazz to experimental and contemporary music, all have in common exigency and demanding research. They often explore the contemporary instrumental and technological limits, bringing new ideas and realizations to the process of creation. Hence musicians and composers contribute heavily to reflection on musical forms and new possibilities of expression. The question of time touches upon technical, musical and philosophical issues, as time meaning or time articulation between the score and the performance. The *Musical time* of improvisations draws a blurred boundary between the notions of writing, composition and transcription. Among our personal collaborations, saxophone player Steve Coleman [[Coleman, Nouno, and Suzuki, 1999](#)], with *Rameses 2000* project in Paris’ Théâtre des Bouffes du Nord, and flute player Malik Mezzadri [[Mezzadri, van der Werf, and Octurn, 2003-2007](#)] raise in their musical quest multiple questions, propositions and desires for the ideal synchronization between computers and musicians. We hope that our work contributes to these goals.

Organization of the thesis

Below, we summarize the following chapters with a few details and references.

Chapter ([1](#)) presents our approach. Temporal interaction between musician and computer is the main purpose of our research and the focus of our observations. We introduce the term *digital ethnology* as a definition of a way to consider the set of interactive relations ([1.2](#)) between computer science and the musician. We explain the reasons and motivations of our work, particularly in the context of real musical experiences

(1.3). We summarize the different definitions and ideas behind the concept of the perception of musical time (1.4), (1.5), (1.6). This leads us towards the concept of rhythm induction and musical anticipation (1.7) that we consider useful for tempo prediction.

Chapter (2) gives a few examples of projects using or needing *tempo follower* for synchronization purposes. We review fundamental psycho-cognitive research on time perception (2.6) and collect those elements and principles that emphasize the hypothesis of our research. We survey in (2.7) the state-of-the-art in *tempo follower* and rhythm induction algorithms. We consider more closely the algorithm of [Large and Jones, 1999], [Large and Palmer, 2002] (5.2) which introduces the important notion of *attending focus* which we will mix with our own research to unwrap the phase of the *tempo* while predicting the pulsation of the *tempo follower* (5.3).

In chapter (3), we propose a mathematical paradigm based on an *optimal grid* search method in a tempo perception assumption framework [Chua and Lu, 2005]. The *optimal grid* can be thought as a regular time-spanned grid of pulsation which best coincides with the observed incoming musical events.

In order to slowly built upon the complexity of the model, we first assume a quasi stable tempo. We define in (3.1) the relations between the regular time-spanned grid that best approximates the input events it is applied to. We propose in (3.1.2) a geometric way to solve this problem, defining the *optimal grid* with parameters s as center and Δ as step. We need to minimize a bidimensional function related to *grid coincidence*, resulting in naming the grid *optimal grid* (3.3). Before doing this, we demonstrate a theorem (Th.3.2.4) which enables us to build an algorithm of complexity $\mathcal{O}(n^3p^2)$ for the optimal search of s and Δ in a continuous two-dimensional region, n is the number of timed input events and p the number of grid points that encompass the events. Theorem (Th.3.2.4) states that, “*considering a temporal regular grid of step Δ and center s which best coincides with the set $\{t_n\}$ of input events, at least one event and a point of the grid occur simultaneously, i.e. are the same point*”. A second *optimal grid* theorem (Th.3.3.9) generalizes the previous result and ensures that “*the optimal grid handles at least two input points* (3.3.5)”.

Chapter (4) adds one level of complexity by considering a continuous and uniform accelerated tempo (4.1). We extend the research of [Guerino Mazzola, 1994] on tempo curves, which brings us to the field of musical transcription. We introduce a simple formulation of two reciprocal operators which handle temporal transformations, the first by uniform direct acceleration (4.1.3) and the second by uniform reciprocal ac-

celeration (4.1.4). We merge these results with the *optimal grid* paradigm, adding a new acceleration parameter θ and ending up with a three-dimensional space function to minimize. We then propose a new iterative method (4.2) to approach numerically the best solution $(\bar{s}, \bar{\Delta}, \bar{\theta})$ and give the final *optimal grid* algorithm of the tempo modeled with a uniform acceleration (4.3).

We show that our paradigm, as emphasized by the results of rhythmic perception studies, invites us to consider perception and rhythm induction as an intrinsic time transcription of events through a quantification performed by the *optimal grid*, as if our cognitive system were organizing itself the perceived events. In (4.4) we draw the conclusion that there is not a unique solution for the transcription, but rather there are several, each handling its own uncertainty leading us to use a *Dynamic Bayesian Network* framework (*DBN*) to solve the *tempo follower* inference.

We present in chapter (5) the dynamic tempo inference, inducted from the estimation method given in chapter (3). In a dynamic temporal context, an observation window of n events represents an abstraction of the short term time memory of the last n events. We define the underlying *optimal grid* of pulses – or *tactus* – as resulting from the tradeoff between the *easiest* or *best* transcribed notation and the *best* tempo variation. The dynamic transcription, or real time decoding tempo inference, is done by the propagation of tempo estimation over time, with the *optimal grid* behaving like a *time quantifier* of the events which are defining the grid. In (5.1.2) we come up with a probabilistic model issued from the observed data, and a sequence of *Hidden Markov Models* (*HMM*). We solve the dynamic inference with a local *Viterbi* decoding algorithm (5.1.3) applied to the hidden states of the tempo. We look at *tempo following* and *tempo phase following* as two separate problems and propose to unwrap the tempo phase with the *attending focus* function [Large and Jones, 1999] as introduced in (5.2). We decompose the *tempo follower* into two steps : 1) tempo inference, and 2) synchronization by entrainment of the chosen inferred grid in response to the external musical events or *stimuli* (5.3). Eventually, we end up with an implementation of our *tempo follower* algorithm in a real time environment (5.4) and discuss some of the results (5.5).

Chapter (6) draws conclusions from the current work and speculates on future research, extensions, and several envisioned contextual applications.

Résumé

à propos des *Musiques Improvisées*

Cette thèse intitulée *Suivi de Tempo Appliqué aux Musiques Improvisées* a pour objet le temps musical pulsé : un temps qui induit une sensation de pulsation et de rythme. Le terme *Musiques Improvisées* peut sembler à la fois trop large ou trop restrictif. L'improvisation évoque une musique spontanée, pleine de surprises pour celui qui écoute comme pour celui qui joue. Elle renvoie également à la connaissance ou non du contexte culturel qui définit son origine, une musique paraît improvisée pour certaines personnes, mais non pour d'autres qui en connaissent ou en saisissent les codes. *Musiques Improvisées* : musiques avant tout exposées, composées ou non. Indépendamment d'un style musical précis, c'est le caractère de création et de génération dans son rapport au temps que nous retenons. Pour l'ordinateur, la "vision" ou "l'écoute" se fait à travers l'acquisition et le traitement d'informations. Nous plaçons notre recherche dans le cadre suivant : l'ordinateur reçoit des informations temporelles musicales pour lesquelles il n'a pas de clef *a priori*, c'est-à-dire que la musique lui apparaît comme une improvisation, un signal informatif généré qu'il observe dans le temps à l'aide de processus algorithmiques.

Tempo musical et suivi de tempo

Le tempo suggère une temporalité particulière. À travers le concept abstrait de *temps musical*, nous nous intéressons au glissement du temps de la *performance*, ou de l'exécution musicale, par rapport à un temps symbolique de la *partition*. Le temps symbolique existe réellement ou non sous une forme d'écriture, mais finalement il génère la *performance*. Cette thèse repose sur la synchronisation d'une temporalité duale. Nous proposons un modèle original d'estimation de tempo basé sur la recherche d'une *grille optimale* (3) : une grille régulière de temps ou de pulsations qui coïncident le mieux possible avec les événements musicaux. Nous réalisons un suivi de tempo (5) dans le cadre d'une inférence dynamique complétée par un modèle de *focus attentionnel* (5.2) issu de la *théorie de l'attention dynamique* de la psychologie cognitive. Nous intégrons notre paradigme de suivi de tempo dans un *environnement musical en temps réel*. Notre intention est d'entretenir dans ce contexte une relation personne-machine pertinente qui puisse nourrir constamment la forme et la pensée complexe de la musique en création.

Plan de lecture de la thèse

Nous résumons ci-après les différents chapitres qui organisent notre recherche.

Le chapitre (1) expose notre approche et notre motivation au contact d'expériences musicales. Nous développons la notion de temps musical dans son rapport avec les outils informatiques.

Le chapitre (2) identifie la problématique du suivi de tempo. Nous donnons un aperçu de l'état de l'art sur le sujet et présentons des éléments de recherche psychocognitive liés à la perception rythmique.

Le chapitre (3) a pour objectif l'estimation du tempo à un instant donné sur une courte fenêtre d'événements musicaux observés jusqu'à cet instant. Nous développons un modèle mathématique de *grille optimale*, la suite d'événements est supposée avoir un tempo quasi constant dans un premier temps. La *grille optimale* est une grille régulière de temps qui coïncide le mieux possible avec les événements observés. Nous démontrons deux théorèmes fondamentaux à partir desquels nous élaborons un algorithme de complexité satisfaisante pour la recherche des *grilles* solutions. Nous considérons par la

suite un tempo accéléré afin de compléter notre modèle et en proposons une nouvelle résolution.

Nous considérons dans le chapitre (5) l'inférence dynamique du tempo, c.-à-d. le suivi de tempo au cours du temps. Nous utilisons un cadre probabiliste dans lequel le tempo est une variable d'état cachée que nous désirons connaître. Nous introduisons un modèle d'oscillateur non linéaire issu de la recherche psychocognitive [Large and Palmer, 2002] pour synchroniser les événements musicaux observés et la pulsation prédite à partir de l'inférence du tempo. Nous présentons une implémentation de notre algorithme de suivi de tempo dans un environnement en temps réel et discutons quelques résultats.

Le chapitre (6) conclut notre thèse, expose des extensions futures sur le plan de l'utilisation et propose d'autres perspectives de recherche à explorer.

Chapitre 1

Introduction

La musique, la science et la technologie entretiennent entre elles depuis longtemps – voire toujours ? – de multiples et fortes relations [Luening, 1968]. Au siècle dernier, la musique et l’informatique ont créé un nouveau contexte dans lequel sont apparues des interactions qui ont ouvert chacune une ou plusieurs branches inexplorées, singulières tant sur le plan musical que sur le plan scientifique et technologique. Ce domaine dénommé sous le terme anglo-saxon assez neutre et générique de *Computer Music*, alimente maintenant depuis plus d’un demi-siècle la recherche artistique et scientifique et convoque pour son développement des espaces qui peuvent sembler a priori très éloignés. Le terme *Computer Music*, traduit sans grande satisfaction par *musique informatique*, *informatique musicale* ou peut-être maintenant de manière trop connotée par le terme de *musique électronique* ne désigne pas un style mais la mise en commun de deux patrimoines. Un patrimoine artistique et un patrimoine scientifique vivants – ou un patrimoine scientifique et un patrimoine artistique vivants [Puckette, 2002]. Il appartient à chacun de décider et d’apprécier à travers son expérience personnelle si le résultat est une science artistique, un art scientifique ou tout simplement un nouveau domaine culturel – et par extension social, économique et politique – qui dépasse de loin les étiquettes qu’on veut lui attribuer. Chadabe [2000] appuie nos remarques dans ce sens.

1.1 Interaction temporelle musique et informatique

Temps, musique et informatique

Parmi les multiples champs d'interactions ouverts par la *Computer Music*, c'est le domaine temporel qui mobilise l'intérêt de notre recherche. Nous nous attachons à une interaction temporelle orientée plus particulièrement sur la question de la synchronisation en temps réel de sources musicales, notamment entre un musicien interprète et un ordinateur. Notre objectif n'est pas de réaliser un suivi de tempo pour une prouesse technologique. Un suivi de tempo peut être trop réducteur et désigne différentes choses suivant les objectifs qu'il propose : il peut s'agir d'une estimation du tempo global d'un morceau de musique, de synchronisation automatique de séquences à tempo constants ou bien de recherche de tempo en vue d'une classification d'une base de données musicale. Nous plaçons notre recherche dans un cadre causal et temps réel. Nous nous intéressons à l'élaboration et à l'intégration d'agents intelligents dans le développement dynamique d'un environnement musical complexe.

Vers un idéal de synchronisation de la machine et du musicien

Pourquoi la synchronisation musicale ? Avant même de s'attacher à des considérations esthétiques – du fait de jouer en même temps, d'être ensemble – la synchronisation vient du désir de communiquer, de partager quelque chose, une émotion, dans la dynamique du temps. La synchronisation musicale a pour origine la nécessité d'une participation active des musiciens, les *ensembles* musicaux en sont l'évidence même. La synchronisation suppose une écoute mutuelle.

De l'avant-garde Jazz aux musiques expérimentales à la musique contemporaine, toutes ont en commun une exigence et une forme de recherche poussée souvent aux limites de l'actuel disponible – instrumental et technologique. Cela stimule bien entendu le développement de nouvelles idées ou réalisations et constitue un moteur fort pour la création. Ainsi naissent les réflexions sur l'élaboration de formes musicales, les possibilités de les contrôler *in situ* ou non. Ainsi sont alimentées les idées de recherche que nourrit également une dimension philosophique sur la question du temps en musique, notamment sur l'articulation du temps de la partition et du temps musical de l'interprétation. Le temps de l'improvisation dessine une frontière mouvante, une

limite dynamique entre les notions d'écriture, de composition, de transcription et d'interprétation. Parmi nos collaborations personnelles, citons le saxophoniste Steve Coleman [Coleman, Nouno, and Suzuki, 1999] – avec le projet *Rameses 2000* au théâtre des Bouffes du Nord, Paris – et le flûtiste Malik Mezzadri [Mezzadri, van der Werf, and Octurn, 2003-2007] qui soulèvent constamment dans leurs quêtes musicales de multiples questions, propositions et désirs sur un idéal de synchronisation de la machine et du musicien. Nous espérons que ce travail apportera quelques pierres à cet édifice.

Formalisation du temps musical pour un paradigme de suivi de tempo

Il nous est apparu au cours de ce travail essentiel de formaliser et d'abstraire les *informations* qui définissent notre approche, afin de les théoriser dans un cadre général. Notre cheminement passe par une formalisation mathématique telle une mise en abîme de la pensée dans la réduction des éléments qu'elle met en œuvre (4.1.4) et (4.2) : une manière de relier les relations macroscopiques complexes de la réalité appréhendée à un modèle qui a pour finalité d'articuler une structure organisatrice sous-jacente. Dans ce contexte dynamique de la musique, nous parlons également d'anticipation rythmique du suivi de tempo qui prédit la pulsation et devient un nouvel acteur dans l'interactivité musicale .

Autour du concept abstrait du temps musical, nous expliquons ci-après comment musique et informatique se rejoignent dans une notion que nous désignons par *ethnologie numérique*.

1.2 Ethnologie numérique

Algorithmie et perception du temps musical

Comment et en quoi un algorithme permet d'appréhender le *temps musical* à travers des hypothèses sur la perception et la compréhension de la sensation de temps organisé provenant d'un ensemble d'événements temporels ? Un flux temporel peut sous certaines conditions devenir un rythme [Grahm and Brett, 2007], [Wright, 2008]. Nous pouvons parler à travers la perception auditive de l'émergence du rythme, notamment dans la propension à créer une synchronisation que l'on peut observer sous des échelles

temporelles diverses [Seebeck et al., 2005]. La perception rythmique est premièrement sensation puis *perception* comprise par le corps avant de l'être par l'esprit conscient. Celui-ci organise les données sensorielles et se forme ainsi une représentation de l'objet extérieur afin de prendre connaissance du réel. C'est dans ce sens qu'on peut parler d'une *ethnologie* de la sensation réinterprétée lors d'un processus algorithmique qui l'approche. Le terme *ethnologie* doit être compris ici comme une tentative de parvenir à l'énonciation de la structure et de l'évolution d'un flux temporel d'informations issues d'un contexte musical, à travers une étude explicative et comparative de l'ensemble de ses caractères ou paramètres. Ce processus de captation du rythme qui, sans vouloir se substituer à notre mécanisme biologique – indéfiniment perfectible – de compréhension des événements extérieurs, reste parallèle à celui-ci dans le rapport des éléments qu'il manipule et transforme, afin d'aboutir à une formulation de la structure temporelle musicale dans des contextes stylistiques précis. Pourquoi ne pas proposer alors une approche *ethnologique* de la *vision* ou de la *perception* rythmique avec l'outil informatique ? Ajoutons alors de façon plus synthétique que l'analyse de notre mode propre de compréhension est nécessaire à l'élaboration d'un algorithme d'estimation et de suivi de tempo, ce qui implique comme nous allons le développer plus loin un *modèle d'estimation* et un *modèle de suivi* du tempo.

Ethnomusicologie computationnelle, ethnomathématique

Le terme *ethnologie*, dans la relation qu'il entretient avec les nouveaux outils numériques appliqués à la musique, est depuis peu utilisé sous le terme de *Computational Ethnomusicology* [Tzanetakis and Kapur, 2007] dans sa référence à la conception, au développement et à l'usage des outils informatiques qui ont un fort potentiel d'utilisation dans l'assistance de la recherche musicologique.

L'ethnomathématique [Chemillier, 2007] met en évidence des structures mathématiques dans certaines activités spécialisées de sociétés de tradition orale. L'une des difficultés qui se posent dans cette approche est de mettre en relation les structures formelles, étudiées "en laboratoire", avec les modes de penser locaux des personnes concernées, tel qu'ils peuvent être observés "sur le terrain". Ainsi, découvrir des organisations et des structures sous-jacentes est établir et mettre à jour des relations qui sont parfois ignorées par les propres acteurs.

Sous la dénomination récente de *MIR*¹, se regroupe avec une certaine approche à l’ethnomusicologie computationnelle, un ensemble d’idées, d’algorithmes, d’outils et de systèmes dans un but d’organisation, de compréhension et de recherche dans d’importantes bases de données musicales. Néanmoins, si la grande variété des problèmes adressés par *MIR* comprend des sujets d’analyse structurelle, d’accompagnement automatique, de suivi de partition ou de suivi de tempo, ils sont pour la plupart trop restrictifs et spécifiques [Parry and Essa, 2003], [Takeda et al., 2003], [Arifi et al., 2003] par rapport à nos objectifs. La raison principale est que l’accent de cette recherche est posé sur une proposition globale qui veut traiter un nombre de cas musicaux [Dixon et al., 2003] statistiquement les plus représentatifs [Miura and Shioya, 2003]. Notre recherche porte sur un contexte musical en temps réel sans a priori de styles, ce qui nous éloigne des objectifs représentatifs de *MIR*.

1.3 Cheminements

L’ethnologie numérique décrite précédemment nous permet de modéliser notre approche et de considérer l’interaction dans son ensemble liée à la perception, malgré une réduction inévitable de la réalité puisqu’il s’agit d’un modèle.

Il y a de réelles difficultés à cerner un si vaste sujet qui paraît pourtant si simple et nature au premier abord : suivre le tempo d’une musique que l’on écoute. Nous nous intéressons à une interaction en rapport avec notre mode de perception analogique ou biologique, nous pensons qu’il y a une nécessité d’inclure un modèle psycho-cognitif de la perception à un modèle d’analyse numérique. Concernant la captation des éléments rythmiques, leurs choix et leurs pré-traitements demandent dans un premier temps une réflexion musicale afin de déterminer les éléments constitutifs de la recherche, de ce qu’on veut leur faire dire ou faire faire. L’originalité de notre approche réside dans l’inspiration la finalité : les sources et les buts viennent principalement des expériences musicales partagées avec les compositeurs et les musiciens, sur le terrain de la création musicale, des studios aux répétitions, des répétitions aux concerts et des concerts aux enregistrements. La dimension pratique est importante afin de respecter une contrainte de faisabilité dans notre application musicale.

Sur l’aspect rythmique, ces discussions ont conduit petit à petit à la lecture d’ou-

¹Music Information Retrieval

vrages d’ethnomusicologie dans la perspective de comprendre pour modéliser, approcher, transcrire ou interpréter les notions nécessaires à la mise en œuvre de la recherche. Il nous était aussi important de lire ces ouvrages pour apporter, confirmer ou refuser des éléments plus ou moins pertinents de notre sujet. Le besoin de fondations claires et précises pour étayer les arguments théoriques a mis à jour des définitions plus solides et confirmé certaines intuitions. Les écrits de Arom [1985a] ont été d’un grand recours, notamment à travers ses études de terrain concernant la pulsation implicite, et pour désigner sans ambiguïté les éléments de la recherche. Nous rappellerons quelques concepts déterminants dans l’analyse rythmique, ainsi que les définitions importantes proposées par l’auteur. Nous nous en ressaisissons dans l’élaboration de la dualité musicien-ordinateur en prenant en compte le *tactus* – la plus petite pulsation pensée – dans la perspective de la *performance* – dans son acception anglophone – en temps réel. Il est intéressant de remarquer que le sens de *performance* en français est lié à un mode d’expression en art contemporain : il consiste à participer à un événement qui comporte généralement une part d’*improvisation* et dont le déroulement même représente l’œuvre.

1.4 Éléments d’analyse musicale, temps et culture

L’observation première que nous considérons est une phase musicale qui porte des informations rythmiques. À travers le flux d’événements provenant de l’exécution musicale, nous pouvons analyser dans le temps plusieurs paramètres comme les hauteurs, attaques, timbre ou divers indices spectraux. Nous voulons décrire une fonction propre à l’interprétation en regard de l’organisation temporelle de la séquence. Nous nous adressons de préférence à un domaine de la musique qui, à travers une forme musicale improvisée ou non et sans considération *a priori* de style, présente à la perception un ou plusieurs cycles temporels prononcés qui induisent des caractéristiques comme que le tempo, la pulsation. Nous verrons s’il est possible d’élargir cette étude à des formes musicales dont l’écriture plus libre et moins définie rend la perception des cycles rythmiques moins évidente. Pour le moment, la *répétition*, la *commutation* sont des critères qui nous aident à formaliser la mise en paradigme. L’estimation de tempo et son suivi en temps réel posent aussi la question de notre compréhension formelle de la polyrythmie, mais établit des niveaux de perception de pulsation. Comment par exemple tenir compte des éléments polyphoniques et du rôle qu’ils jouent à l’écoute

de leurs modes d'imbrication les uns avec les autres ?

D'un point de vue expérimental, nous essaierons de capter de manière séparée les événements sonores issus d'un même instrument ou ensemble d'instruments. Chacun construit un flux temporel rythmique élémentaire. Le flux rythmique global résulte de la totalité et par extension de *l'inter-action* ou de la conjonction des flux rythmiques élémentaires. Néanmoins, il n'est pas toujours possible d'avoir accès de manière exhaustive à tous les flux rythmiques élémentaires : soit pour des raisons pratiques, soit parce qu'ils ne sont pas nécessairement exprimés ou joués par les interprètes. Pourquoi donc ? Un flux rythmique peut être caché ou sous entendu de par la culture musicale que partagent les interprètes. Il est aussi difficile de séparer ce qui est culturel de ce qui ne l'est pas. Traiter une information de manière isolée, voire décontextualisée, est réducteur mais permet parfois de trouver les éléments minimum de reconstruction d'une information musicale, même partielle. Des structures rythmiques identiques peuvent être perçues différemment chez des musiciens de culture différente, car les repères, les référents culturels ne sont pas les mêmes. Parfois, les interprètes jouent sur une pulsation commune qui n'est pas exprimée ou bien qui est exprimée partiellement, avec peu de matérialisation musicale ce qui la rend moins évidente voire invisible. Il y a là presque un paradoxe d'entendre des musiciens jouer ensemble mais de ne pas sentir ce qui les fait jouer ensemble. C'est une question fondamentale à laquelle nous essaierons d'apporter un élément de réponse avec la recherche de la grille optimale exposée au chapitre (3). Ainsi nous intéressons-nous à la forme car, pour reprendre les mots de Arom [1985a] à propos de ses travaux d'ethnomusicologue en Afrique : "l'objet de cette étude s'appuie sur un système purement formel qu'est la musique dépourvue d'épaisseur sémantique".

1.5 Tempo, pulsation, tactus et tatum

Nous sommes amenés implicitement à nous intéresser aux musiques mesurées, c'est-à-dire constituées de durées dont les valeurs sont proportionnelles. Les durées proportionnelles sont basées sur un repère, un étalon de temps. Les Grecs antiques utilisaient le temps premier, unité de durée minimale et indivisible. En occident, cet étalon est le temps battu par le chef d'orchestre, et le terme italien *a tempo* ou *a batuta* se rapporte au terme mesuré, indiquant que la mesure temporelle découle de la battue. Les notions de temps et de battue se rejoignent dans le contexte musical : la battue

marque le temps, elle indique le tempo. Jusqu'à l'époque de l'*Ars Nova*, l'exécution des polyphonies les plus complexes comme celles de Guillaume de Machaut nécessitaient seulement un simple étalon de temps qui assurait le synchronisme des parties lors de l'exécution et simultanément indiquait le tempo. Cet élément qui préfigure la *batuta* du chef d'orchestre, avait pour nom *tactus* ou littéralement *touchement*.

Les musiques africaines répondent à un principe identique à celui de *tactus* [Arom, 1985b], et la notion de réitération régulière de temps accentués, opposés à d'autres qui ne le seraient pas leur est totalement étrangère. C'est à dire que les musiques africaines ne sont pas fondées sur des mesures mais sur des pulsations : une succession d'unité de temps isochrones pouvant être matérialisées par une battue. Le Jazz et les musiques improvisées héritent en grande partie des racines des musiques africaines et répondent aussi à un principe de pulsation.

Il est intéressant de comprendre comment a évolué le concept de l'organisation du temps dans la polyphonie occidentale. Dans la conception ancienne de la battue, au Moyen Âge, on ne groupe pas les temps, chaque temps est marqué individuellement en une séquence de coups de doigts. Son emploi est demeuré prépondérant jusque 1625, à partir de ce moment un autre mode de mesure, hiérarchisé, s'est introduit : la mesure, *“Le seul fait de grouper les temps en mesure ne fut possible que lorsque la notion de mesure elle-même, notée graphiquement par des barres, envahit le solfège du XVII^e siècle”*. Ainsi la notion de mesure, issue d'une simple convention graphique, exerce toujours une emprise considérable sur la musique occidentale. Ce n'est qu'à partir du XX^e siècle que certains compositeurs commencent à se libérer de son emprise et redécouvrent d'un oeil nouveau les polyphonies anciennes [Nono, 1993] [Maderna and Verzina, 2003].

1.6 Continuité, discontinuité : une arche du temps musical

Imberty [2005] affirme que toutes les valeurs rythmiques peuvent pratiquement se réduire soit à une pulsation unique et régulière, soit à deux pulsations inégales observant la proportion de deux à trois. La psychologie cognitive a depuis montré que ce rapport ou binaire ou ternaire est une donnée biologique et neurophysiologique. Les travaux de Olivetti-Belardinelli [1993] l'ont établi de façon définitive, indiquant en outre

que certains individus sont “naturellement binaire”, produisent des rythmes spontanés de structure binaire, et d’autres “naturellement ternaires”. Olivetti-Belardinelli and Pessa [1991] ont en outre montré que cette prédominance correspond aussi à deux styles cognitifs différents [Olivetti-Belardinelli et al., 1993] [Olivetti-Belardinelli and Nicole, 1993].

Boulez [1963] rappelle que le temps musical peut se décrire en deux catégories : le temps pulsé qu’il appelle le temps strié, et le temps amorphe qu’il appelle le temps lisse. *“Dans le temps pulsé, les structures de durée se référeront au temps chronométrique en fonction d’un repérage, d’un balisage - pourrait-on dire - régulier ou irrégulier, mais systématique : la pulsation, celle-ci étant l’unité la plus petite (plus petit commun multiple de toutes les valeurs utilisées), ou un multiple simple de cette unité (deux ou trois fois sa valeur). Le temps amorphe ne se réfère au temps chronométrique que d’une façon globale ; les durées, avec des proportions (non des valeurs) déterminées ou sans aucune indication de proportion, se manifestent dans un champs de temps. Seul le temps pulsé est susceptible d’être agi par la vitesse, accélération ou décélération ... Le temps amorphe sera seulement plus ou moins dense suivant le nombre statistique d’événements qui arriveront pendant un temps global chronométrique”.*

Comme le souligne l’auteur, à propos du prélude de Parsifal, l’alternance de deux pulsations inégales observant la proportion de deux à trois *“tire le temps musical globalement vers un temps strié directionnel”*. Mais d’autre part, il insiste sur cette organisation régulière de l’irrégulier : *“c’est parce que Boulez se méfie de la tendance à diluer le rythme et la pulsation par les interprètes qui l’ont précédé et qui ont, selon lui, englué la musique dans une continuité amorphe et pesante sous les mauvais prétextes de tradition et de mysticisme”*. On peut parler d’une arche du temps musical qui va de l’arythmicité ou le continu du temps lisse à la pulsation du temps strié.

1.7 Anticipation musicale

Dans ce temps strié, le suivi de tempo, sous un aspect d’informations abstraites, a pour domaine d’étude une séquence de points temporels, manifestés par des événements qui sont des impulsions ou des attaques de notes. Deux tâches se recouvrent, également mises en évidence par des études neurophysiologiques [Cojan, 2006] : une tâche de perception et une tâche de production ou de continuation. La perception manipule la

séquence temporelle rétrospectivement, tandis que la production l'utilise de manière prospective et causale [Pearl, 1985] : on s'intéresse aux événements passés et on désire prévoir au mieux où tombe le prochain événement. L'interaction de la perception et de la prospection n'est pas à sens unique. La nature de l'écoute modifie la prédiction et la nature de la prédiction modifie à son tour la nature de l'écoute. Cette récursivité apparaît comme un élément déterminant de la notion d'anticipation. L'anticipation musicale se définit comme une forme de “croyance” mentale ou corporelle de l'apparition d'un événement dans un futur proche [Huron, 2006]. Mais l'anticipation est plus qu'une simple prédiction, elle est une modification intrinsèque de son comportement même, sur ce qu'elle observe et sur elle-même. Dubnov [2006] décrit un modèle basé sur la théorie de l'information qui permet de quantifier dans un signal audio et musical le degré d'anticipation, comme un indice de la nouveauté. L'auteur introduit une nouvelle approche d'analyse musicale qui implique un profil d'anticipation. Il s'agit là d'établir dans un contexte dynamique la relation de l'aléatoire et du structuré. La réduction du degré d'incertitude sur le signal constitue une forme de mesure mathématique de l'anticipation dans laquelle le signal écouté et l'auditeur sont partie prenante du paradigme. Le degré de structure est relatif à la quantité d'information transmise du passé vers le présent, et dépend à la fois de la nature du signal et de l'auditeur. Cette approche a des conséquences importantes sur la compréhension de systèmes complexes et sur leur modélisation. De nouvelles applications en recherche musicale apparaissent, comme le suivi de partition *Antescofo* [Cont, 2008a],[Cont, 2008b]. *Antescofo* intègre un modèle mathématique de l'anticipation dans la collaboration dynamique d'agents intelligents qui concourent ensemble au décodage d'informations pour réaliser un suivi de partition très robuste .

1.8 Plan de lecture de la thèse

Nous présentons ci-dessous un aperçu comprenant quelques détails et références des thèmes exposés dans les chapitres suivants afin de faciliter la lecture du document.

Chapitre (2)

Dans le chapitre (2), après avoir fait le lien avec les recherches fondamentales psychocognitives sur la perception du temps (2.6), nous présentons un aperçu de l'état de l'art

sur le suivi de tempo et les algorithmes d'induction rythmique (2.7). Nous décrivons plus en détail l'algorithme de [Large and Jones, 1999], [Large and Palmer, 2002] (5.2), qui introduit la notion importante de *focus attentionnel* que nous reprendrons par la suite dans notre recherche (5.3)

Chapitre (3)

Dans le chapitre (3), nous développons un modèle mathématique de *grille optimale* dans le cadre d'une hypothèse perceptive du tempo [Chua and Lu, 2005]. Afin d'avancer progressivement dans la complexité de l'élaboration du modèle, nous nous plaçons dans un premier temps dans le cadre d'un tempo relativement stable. Nous définissons alors en (3.1) la relation d'une grille temporelle régulière quant à l'approximation qu'elle réalise par rapport aux événements auxquels elle se rapporte. Nous proposons en (3.1.2) une méthode de résolution géométrique de la grille que nous caractérisons par son centre s et son pas Δ . Nous abordons ainsi la résolution d'un problème de minimisation d'une fonction bidimensionnelle de coïncidence de la grille que nous appelons en conséquence *grille optimale* (?). Nous avons au préalable démontré un théorème (Th.3.2.4) qui nous permet d'élaborer un algorithme de complexité $\mathcal{O}(n^3p^2)$ pour la recherche optimale de s et Δ , n représente le nombre d'événements donnés et p le nombre de points de grille. Le théorème (Th.3.2.4) énonce que si l'on considère une grille temporelle régulière – métronomique – de pas Δ et de centre s qui coïncide au mieux avec un ensemble d'événements $\{t_n\}$ donnés, alors au moins un événement donné et un point de la grille sont simultanés, i.e. confondus. Un second théorème (Th.3.3.9) sur la *grille optimale* généralise par la suite le résultat précédent et affirme que la *grille optimale* passe par au moins deux points événements donnés, (3.3.5). Nous complexifions notre modèle et abordons alors la problématique d'un tempo uniformément et continûment accéléré (4.1). Nous étendons pour cela des recherches de Guerino Mazzola [1994] sur la notion de modèle de courbe de tempo, ce qui nous renvoie à la notion de transcription musicale. Nous introduisons alors une formulation simple de deux opérateurs caractérisant une transformation temporelle par accélération directe uniforme (4.1.3) et inversement une transformation temporelle par accélération uniforme réciproque (4.1.4). Nous intégrons alors ces résultats dans le modèle de *grille optimale bidimensionnelle* où un nouveau paramètre d'accélération θ entraîne la recherche d'un minimum dans un espace tridimensionnel. Nous développons alors en (4.2) une nouvelle méthode pour approcher la meilleure solution (s, Δ, θ) et donnons, en (4.3), l'algorithme final de

la *grille optimale* d'un tempo modélisé par une accélération uniforme. Nous montrons alors que notre paradigme, appuyé par les recherches de perception du rythme, nous invite à considérer la perception de l'induction rythmique comme une transcription intrinsèque des événements à travers une quantification que réalise la *grille optimale*, notre système cognitif agissant comme auto-organisateur des événements perçus. Nous concluons en (4.4) qu'il n'existe pas une seule solution de transcription, mais plusieurs et que chacune se rapporte à un degré de certitude ou de probabilité qui nous mène vers un cadre de résolution dynamique bayésien pour effectuer le suivi le tempo.

Chapitre (5)

Nous présentons au long du chapitre (5) l'inférence dynamique du tempo induit à partir de son estimation développée dans le chapitre (3). Dans une situation temporelle dynamique, une fenêtre de n événements observés représente une abstraction de la mémoire musicale rythmique à court terme des n derniers événements. Nous définissons la *grille optimale* sous-jacente des pulsations – *tactus* – comme un compromis de la notation ou transcription la plus facile et de la meilleure variation du tempo. La transcription dynamique, ou le décodage du suivi de tempo, s'effectue à partir d'une propagation de l'estimation de tempo et la *grille optimale* agit en tant que quantificatrice des événements qui la créent. Nous présentons en (5.1.2) le modèle probabiliste issu de l'observation de nos données et la séquence par modèle de Markov caché (*HMM*) qui en résulte. En (5.1.3) nous effectuons l'inférence dynamique avec un algorithme de décodage *Viterbi* appliqué à la séquence d'états cachés du tempo. Nous séparons suivi de tempo et suivi de phase du tempo et choisissons de la reconstruire à l'aide de la fonction de *focus attentionnel* introduite en (5.2). Nous décomposons ainsi le suivi de tempo en deux temps : une inférence du tempo et une synchronisation réalisée par un entraînement de la grille de quantification inférée en réponse aux stimulations extérieures des événements (5.3). Nous présentons une implémentation du suivi de tempo en temps réel en (5.4) et discutons ensuite quelques résultats (5.5).

Chapitre (6)

Le chapitre (6) clos notre thèse sur quelques conclusions et extensions futures sur le plan de l'utilisation et des perspectives de recherche à explorer.

Chapitre 2

Suivi de tempo : fondements et état de l’art

Nous présentons ici un aperçu des recherches et technologies les plus influentes concernant le domaine du suivi de tempo et des notions qui lui sont voisines. Les éléments de recherche fondamentale, notamment de perception et de psychologie cognitive, sont difficilement séparables des recherches musicales appliquées, car ces dernières s’appuient nécessairement sur les premières. Nous examinons en (2.3) les différentes données, observables ou non, impliquées dans notre problématique. En (2.6) nous prenons comme points de repère des études psychocognitives importantes pour les hypothèses qu’elles posent et mettent en œuvre et qui sont à la base de nombreuses recherches et réalisations dans les domaines s’appuyant sur la perception du temps musical. Nous évoquons en (2.1) des applications musicales utilisant des interactions rythmiques importantes. Nous faisons référence en (2.7) aux travaux, algorithmes et outils principaux existants dans les domaines proches désignés de façon générique par le terme *suivi de tempo*. Nous décrivons plus en détail en (5.2) l’algorithme de [Large and Jones, 1999], puis développé dans [Large and Palmer, 2002], qui introduit la notion importante de *focus attentionnel* que nous reprendrons par la suite dans notre recherche (5.3).

2.1 Applications musicales

Une part importante de la recherche en informatique musicale provient du désir de reproduire ou de s'approcher de la capacité humaine dans la production et la perception d'événements sonores synchronisés. On trouve de nombreuses applications proposant des systèmes de transcription automatique, des programmes d'accompagnement musicaux à *l'écoute* du musicien, des *partenaires* improvisateurs électroniques.

Lewis [2000] propose *Voyager*, un environnement d'improvisation. Le saxophoniste de jazz Coleman et al. avec le projet *Rameses 2000* écrit dans l'environnement *MaxMsp*¹, définit un environnement d'improvisation dont la logique s'inspire de son mode de pensée. L'ordinateur intervient par analyse des phrases mélodiques, des accords joués et du suivi rythmique du batteur. Le suivi de tempo utilise sur la batterie des capteurs qui transmettent les informations d'attaques instrumentales. Une préversion de l'algorithme de Large [1994] – qui ne comprenait pas encore la notion de *focus attentionnel* décrite plus tard dans [Large and Palmer, 2002] – effectue le décodage du tempo et permet la synchronisation des événements et propositions musicales que l'ordinateur renvoie aux musiciens.

2.2 Quelques définitions

Nous sommes amené à utiliser des termes différents pour désigner des notions souvent très proches, nous précisons ici les équivalences utilisées afin de lever les confusions qui pourraient apparaître à leur lecture.

Tactus, pulsation, “beat”, temps, battue

“Ces divers termes désignent, dans toute musique mesurée, l'élément d'une série de repères régulièrement espacés, i.e. équidistants dans le temps, qui délimitent le continuum musical en unités égales, et servent ainsi d'étalon pour l'organisation de toutes les durées qui y figurent” [Arom, 1985a]. Nous regardons dans notre recherche des séquences d'événements dans le temps pour lesquelles nous ne prenons pas en compte

¹<http://www.cycling74.com>

la hiérarchie métrique et ne considérons pas les polymétries. Nous utiliserons le terme de *pulsation* ou de *tactus* pour désigner la plus petite pulsation étalon de la phrase rythmique analysée. Nous utiliserons parfois de manière équivalente la notion de *temps* ou de “*beat*” pour les références aux textes de langue anglaise.

Suivi de tempo

Nous parlerons de *suivi de tempo* en faisant référence parfois aux termes anglophones “*tempo tracking*”, “*beat tracking*”, “*beat follower*”, “*tempo induction*” et “*beat induction*”. L’expression *suivi de tempo* présuppose un caractère dynamique, entendu également dans *tracking* ou *follower*, par opposition à *estimation de tempo* qui présente un aspect plus statique, à un instant donné, ou global, pour toute la durée d’une séquence par exemple. La distinction entre *suivi* et *estimation* n’est cependant pas toujours claire dans les textes. Comme nous le définissons ci-après, le comportement ou caractère du *suivi de tempo* n’est pas implicitement donné par sa seule dénomination.

2.3 Identification des problématiques

Le rythme, tout comme l’harmonie, la mélodie et le timbre, représente un aspect fondamental de la musique. Ce caractère fondamental évident et naturel est paradoxalement difficile à expliciter. Le rythme musical est une perception qui émerge de l’organisation temporelle des éléments sonores, il peut être aussi visuel dans d’autres domaines artistiques. Le rythme implique intrinsèquement une forme cyclique, ou du moins une répétition ou des répétitions avec variations. Par analogie, on peut voir des répétitions et variations rythmiques dans des constructions architecturales – notamment chez Le Corbusier [Loach, 1998] et Frank Lloyd Wright [Turner, 1983] – ou dans certaines oeuvres d’art plastique – Piet Mondrian [Bois and Reiter-McIntosh, 1988]. Percevoir un rythme est percevoir une forme [Wallin et al., 2000], peut-être pourrait-on avancer que la perception rythmique est une *interprétation* ou une *transcription* d’une forme.

Le terme *suivi de tempo* indique l’action de suivre le tempo d’un flux d’événements dans le temps, mais suivant la nature des données observées et résultantes il peut être désigné autrement. Les premières recherches sur le *suivi de tempo* remontent aux années 1980 et portaient principalement sur la validation de modèles psychologiques

de la perception rythmique. Avec l'évolution des capacités de calcul, les travaux de recherche se sont progressivement déplacés des domaines de la perception musicale vers les sciences informatiques et vers une ingénierie algorithmique appliquées. On le voit notamment parmi les sujets du domaine récent de *Music Information Retrieval* (*MIR*)², qui substitue souvent un problème de classification à un problème de perception [Wells, 2003],[Regina, 2006]. Nous y trouvons des thèmes émergents fortement liés à la complexité de l'information musicale [Downie, 2004] mais principalement appliqués à un corpus musical réduit, commercial et stylistiquement très stéréotypé – *western music*.

2.4 Prétraitement du signal musical

Nous désirons réduire le domaine d'étude et d'observation au minimum indispensable pour aborder le problème de suivi de tempo dans le cadre le plus général. Comme nous le précisons par la suite, nous considérons la *première hypothèse* suivante tout au long de notre travail : nous observons un signal musical préalablement segmenté dans le temps. Nous discutons ci-dessous de la segmentation comme un prétraitement du signal.

De l'information musicale à l'infomation symbolique

Le déplacement des recherches de perception sur le tempo vers un aspect plus appliqué nécessite un prétraitement de l'information musicale : il faut passer d'une séquence ou phrase musicale analogique à un signal symbolique sur lequel travailler. Notre appréhension du temps musical et du tempo passe par une phase d'abstraction. C'est aux recherches en psychologie cognitive de la mettre en évidence, et en conséquence de fournir des éléments pour intégrer cette phase dans les traitements informatiques. La figure Fig.2.1 présente un schéma global du traitement de l'information musicale : depuis son émission, jusqu'à son traitement, analyse et retour vers le musicien sous la forme d'une intervention musicale active. L'objectif principal du prétraitement en vue d'un suivi de tempo est de réaliser une segmentation du signal musical.

²<http://www.ismir.net/>

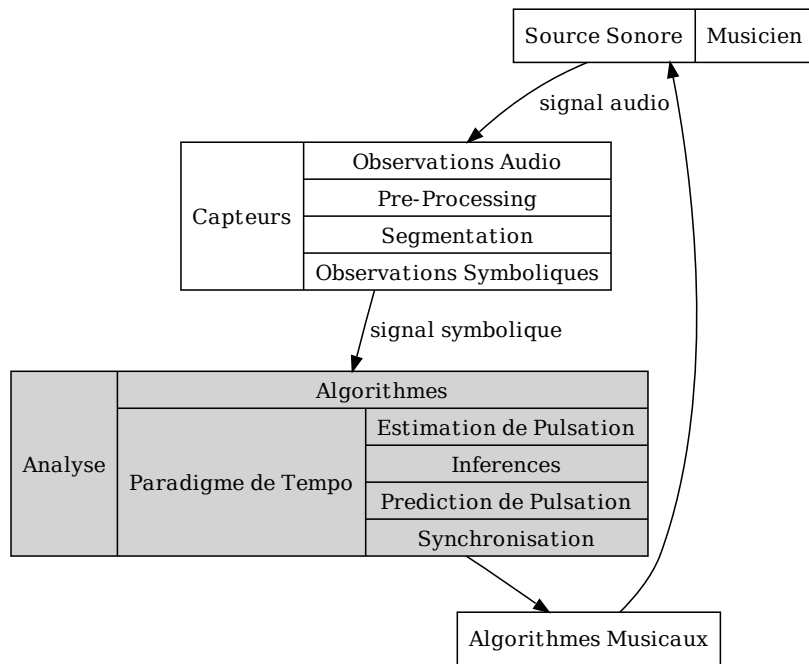


FIG. 2.1 – Schéma général du suivi de tempo, de la source au suivi

Nature du signal musical

Les données observées par un suivi de tempo et les données résultantes en déterminent l'approche et les objectifs, cela nous permet de caractériser les différents types de suivi de tempo. Nous pouvons proposer une catégorisation ou une classification des différentes démarches à partir des informations en entrée et en sortie. La nature des données observées en entrée d'un système de suivi de tempo est soit un signal symbolique soit un signal audio. S'il est vrai que, dans un mode d'écoute, le signal acoustique est préexistant au symbolique, il y a toujours pour le suivi de tempo une étape d'abstraction qui transforme le signal audio en données symboliques. À un endroit du processus d'analyse est nécessaire une segmentation qui détermine une classe d'événement ou un état symbolique associé à un découpage du signal sonore – note, accord, attaque, bruit ...

La perception du tempo est liée à la reconnaissance d'une régularité [Dixon et al., 2004], d'une forme qui se répète [HOOPEN et al., 2006], d'où une opération de représentation symbolique des données vers un espace d'états afin de décrire la forme. La distinction symbolique ou audio de l'entrée n'est alors pas si fondamentale, car l'étape de segmentation est toujours présente et c'est elle qui conditionne le fonctionnement de l'algo-

rythme d'estimation et de suivi de tempo en aval. On peut voir la segmentation comme un processus en amont des données d'entrée – une étape très importante dans un processus de reconnaissance [Duda et al., 2000], comme une extraction de descripteurs – *features* – sur lesquels on travaille par la suite. Ainsi, on peut séparer la problématique de l'induction rythmique en deux phases distinctes : 1) la détermination d'attaque acoustique [Wright, 2008] qui induit une pulsation, et 2) la recherche d'une périodicité qui induit le rythme et le tempo [Gouyon et al., 2006]. Notons qu'une périodicité maximale entraîne une estimation de la métrique ou d'un *pattern* rythmique alors qu'une périodicité minimale suppose la recherche du *tactus* qui nous intéresse. En d'autres mots, la distinction des algorithmes qui travaillent soit sur le signal, soit sur un signal déjà segmenté n'est pas très pertinente, ce qui est plus pertinent est comment la segmentation est effectuée [Mierswa and Morik, 2005] et quel est l'espace des variables d'états retenu, augmenté ou non d'autres variables cachées.

2.5 Variables observables et non observables, hypothèses pour un suivi de tempo temps réel

Comme nous venons de le préciser ci-dessus, la segmentation suivant un critère d'attaque – ou *onset* – ramène le signal audio et le signal symbolique sur un même plan : un signal d'information symbolique, car découpé. À la place de la distinction audio/symbolique, nous faisons la distinction sur les variables que l'observation induit à partir de l'observation du nouveau signal d'information. Ces variables peuvent être observables ou cachées, connues ou inconnues. Dans un signal musical, nous pouvons observer la hauteur, l'intensité, les attaques et le timbre ainsi que d'autres descripteurs. Nous choisissons de ne pas nous intéresser aux paramètres de hauteur et d'intensité, non pas parce qu'ils ne sont pas intéressants, mais par souci de simplification du problème. Nous projetons le signal musical sur l'axe du temps pour le réduire à cette seule dimension que nous supposons essentielle et suffisante pour le suivi de tempo. Nous posons ici une *seconde hypothèse* forte : l'unicité de la dimension revient à décrire un rythme comme constitué uniquement d'oppositions de durées, et à mettre de côté les oppositions d'intensité et de timbre. Ce faisant, nous ne prenons pas en considération la métrique fondée sur les oppositions d'accents – les temps faibles et les temps forts caractéristiques des mesures.

Prenons un exemple et considérons le jeu d'un musicien : on peut abstraire le flux d'informations acoustiques généré en une séquence de points temporels qui a pour manifestation un ensemble d'événements musicaux représentés par des attaques ou des impulsions. La segmentation de la phrase peut être réalisée de plusieurs façons, en modélisant la perception des attaques [Wright, 2008] ou bien avec des techniques mathématiques faisant appel à la théorie de l'information [Cont, 2008b] qui élargit le concept d'attaque en considérant une mesure de la nouveauté perçue [Dubnov, 2006]. Les données observables de la performance musicale sont les événements que l'auditeur et le musicien écoutent, indépendamment de qui participe à leur production. Le tempo auquel joue le musicien n'est quant à lui pas directement observable, nous le considérons comme une variable cachée. Quand le musicien joue la séquence d'événements, il pense probablement à la représentation symbolique de ce qu'il est en train de jouer. Cette représentation est parfois corporelle : inscrite dans le corps et matérialisée par un mouvement du musicien – ou du danseur – [Paczynski, 2000].

Une représentation symbolique de la musique jouée est soit connue, si le musicien est en train de lire une *partition*, et nous passons ici dans une problématique de suivi de partition, ou bien inconnue : le musicien n'a pas de partition ou improvise. Pour l'auditeur, les mêmes remarques tiennent, il peut connaître la *partition* des événements musicaux ou non. Notre centre d'attention est la performance en temps réel pour laquelle nous n'avons pas accès à la *partition* des événements, le musicien est libre de jouer ce qu'il veut. La *troisième hypothèse* que nous formulons est qu'il existe dans la séquence musicale jouée un minimum d'organisation temporelle qui permette une induction du tempo [Gouyon et al., 2006].

En résumé, les trois hypothèses que nous posons pour un suivi de tempo dans un cadre général en temps réel sont les suivantes :

1. nous observons un signal musical préalablement segmenté dans le temps.
2. les oppositions de durées sont suffisantes pour créer un rythme, seule la dimension du temps est retenue.
3. l'écoute de la séquence musicale permet d'induire un tempo

2.6 Perception du temps musical : études psychocognitives

Quand nous écoutons un rythme, nous dégageons spontanément des signes, gestes ou attitudes d'évidence de ce rythme. [Grahn and Brett \[2007\]](#) montrent que ce phénomène témoigne certainement du traitement du rythme ou du tempo par la région moteur du cerveau. L'induction du tempo par un rythme a été mise en évidence de manière claire avec des rythmes dont les intervalles entretiennent des relations entières et des accents perceptifs réguliers. Du point de vue physiologique, il apparaît récemment très probable que des régions localisées du cerveau – SMAs et ganglia – soient les vecteurs de la perception de la pulsation ou *beat*, et que la région moteur du cerveau soit active dans ce processus [[Chen et al., 2008](#)].

Théorie de l'attention dynamique

Malgré sa complexité, la musique propose souvent une écoute cohérente et un certain degré de prédiction qui nous font dire que les événements n'arrivent que rarement au hasard et semblent obéir à certaines règles. Il est important de distinguer ici la prédiction statistique – dans un environnement donné, certains événements sont plus à même de se produire – de la prédiction structurelle – à un moment donné, certains événements sont plus à même de se produire du fait de la configuration structurelle précédente.

Dans le domaine de la perception et de la psychologie cognitive, [Drake, Jones, and Baruch \[2000\]](#) étendent la théorie de l'attention dynamique [[Jones, 1976](#)]. Les auteurs développent les notions de tempo et de hiérarchie temporelle à partir de la relativité dynamique des intervalles de temps, plus grands ou plus petits durant l'écoute, ce qui entraîne une focalisation de l'*attention*. L'*attention* – ou *écoute attentionnelle* – est déterminée par le comportement de *rythmes attentionnels* internes : ce sont des rythmes présents et en attente de synchronisation modélisés par des oscillateurs adaptatifs internes. Il s'en suit une modification de la perception et de la mémoire. Un oscillateur est ici synonyme de *rythme attentionnel* et fait référence à la régularité périodique, représentée par une distribution temporelle de *l'énergie attentionnelle*. L'oscillateur est adaptatif car il ajuste ses paramètres de manière à se synchroniser

avec des périodes de temps extérieures en guidant son attention sur le déroulement des événements.

[Large and Jones \[1999\]](#) concentrent leur recherche sur deux aspects principaux : le tempo et la hiérarchie temporelle des événements. Le tempo est alors modélisé par un seul oscillateur adaptatif dont la période agit comme référence : un niveau hiérarchique de tempo. Un oscillateur interne est supposé auto-entretenu, dans un système dynamique il acquiert une période ou un cycle stable appelé cycle limite. Quand l'oscillateur est couplé à un rythme externe, c'est-à-dire qu'il observe un rythme externe, son oscillation auto-entretenu peut se synchroniser en adaptant sa phase et sa période. Alors que la *période de référence* de l'oscillateur correspond au tempo subjectif perçu, un *niveau de référence* [\[Drake, Jones, and Baruch, 2000\]](#) correspond à un stimulus saillant auquel l'attention se porte dans la séquence d'événements. Autrement dit en terme musical, ce stimulus est le *tactus*, il détermine *quand* l'attention devrait être la plus forte dans le temps. Le terme *tactus* fait lui même référence à l'action de battre le temps avec le doigt, le touché [\[Arom, 1985a\]](#). On désigne par harmonisation l'action par laquelle l'*oscillateur attentionnel* se synchronise avec un stimulus périodique. Ainsi, l'harmonisation consiste à adapter la période de l'oscillateur à un *tactus*. D'un point de vue formel, il s'agit dans cette action de réduire une disparité de phase : temps entre le moment de la *pulsation attentionnelle* portée par l'oscillateur et le moment où le stimulus musical a lieu. Un état d'attraction de phase correspond à une différence de phase nulle entre ces points, un couplage de phase décrit le processus attractif de l'oscillateur non linéaire [\[Large and Kolen, 1994\]](#). Nous utiliserons en (5.2) ce modèle pour la synchronisation de la phase du tempo en combinaison avec la théorie de notre modèle de tempo.

2.7 Etat de l'art

Nous présentons ci-après un état de l'art des techniques les plus influentes liées à l'estimation de tempo et à son suivi, désigné la plupart du temps dans la littérature sous les termes de *beat tracking* et de *tempo follower*. Ces méthodes ont pour but d'estimer le tempo d'un flux d'événements musical et de synchroniser une pulsation avec le rythme suivi – *beat tracking*. [Hainsworth \[2004\]](#) et [Collins \[2004\]](#) présentent un panel relativement détaillé des différentes approches existantes, tandis que [Temperley \[2004\]](#) en propose une évaluation méthodologique, ainsi que [McKinney et al. \[2007\]](#)

pour un nombre réduit d'algorithmes. Les différences essentielles des modèles portent sur la causalité ou non des systèmes, leurs utilisations en temps réel, les buts qu'ils proposent et le type de données employées en entrée, symbolique ou signal audio. Les anciennes méthodes utilisaient des entrées symboliques générées à partir de partition. Les méthodes actuelles utilisent soit une entrée midi – une segmentation évidente – ou un signal audio. Suivant les entrées et les buts de chaque méthode, on peut répartir les algorithmes suivant les axes principaux suivants :

1. jeu de règles
2. autocorrélation
3. filtres oscillants
4. histogrammes
5. agents multiples
6. statistiques

La table [2.1](#) présente un panel des algorithmes principaux présentés par ordre chronologique.

TAB. 2.1 – *Beat Tracking* : algorithmes principaux

référence	mode	remarques
Longuet-Higgins [1979]	approche par règle	musique classique
Allen and Dannenberg [1990]	agents multiples	
Rosenthal [1992]	agents multiples	parsing
Brown [1993]	auto-corrélation	temps réel
Parncutt [1994]	approche par règle	perception
Large [1995]	oscillateur non linéaire	temps réel
McAuley [1995]	oscillateur non linéaire	temps réel
Scheirer [1998]	banque de filtres	temps réel
Toiviainen [1998]	oscillateur non linéaire	temps réel
Temperley and Sleator [1999]	approche par règle	harmonie et rythme
Cemgil et al. [2000] Cemgil [2001]	probabiliste	
Smith [2000]	banque de filtres, ondelettes	
Dixon [2001]	agents multiples	non causal
Eck [2001a]	approche par règle	basé sur les accents
Eck [2001b]	oscillateur non linéaire	temps réel
Foote and Uchihashi [2001]	auto-corrélation	non causal
Goto [2001]	agents multiples	temps réel
Gouyon and Herrera [2001]	histogramme	tick (tatum)
Laroche [2001]	probabiliste	max likelihood
Raphael [2001]	probabiliste	
Seppanen [2001]	histogramme	tatum
Wang and Viterbo [2001]	histogramme	tatum sur mp3
Paulus and Klapuri [2002]	auto-corrélation	non causal, similarité
Tzanetakis [2002]	auto-corrélation	non causal, ondelettes
Jensen and Andersen [2003]	histogramme	temps réel
Klapuri [2003]	probabiliste	
Uhle et al. [2003]	histogramme	+ auto-corrélation
Sethares and Morris [2005]	probabiliste	
Chua and Lu [2005]	auto-corrélation	tempo perceptif
Alonso et al. [2007]	auto-corrélation	
Ellis [2007]	auto-corrélation	non causal
Peeters [2007]	auto-corrélation	onset énergie spectrale
Holzapfel and Stylianou [2008]	probabiliste	Group Delay, HMM

Chapitre 3

Grille optimale : un modèle pour l'estimation de tempo

Dans le présent chapitre, nous développons un modèle mathématique de grille optimale dans le cadre d'une hypothèse de perception d'un tempo quasi stable [Chua and Lu, 2005]. La grille optimale est une séquence mono-dimensionnelle de temps ou de pulsations régulièrement espacées qui coïncident le mieux possible avec les événements musicaux observés. En d'autres termes, la grille optimale est une suite de pulsations homorythmiques dont l'intervalle de temps étalon permet de définir le tempo le plus juste de la séquence musicale donnée.

La notion d'événement qui tombe avant ou après le temps – *beat* – permet à la structure de pulsation régulière une existence plus indépendante qu'on ne le suppose généralement. Une grille bien qu'étant une structure abstraite peut avoir une certaine flexibilité d'étirement, elle représente une construction psychologique plus robuste qui permet néanmoins de maintenir un certain degré d'autonomie des événements [Dixon et al., 2004]. L'interview du contrebassiste Riccardo del Fra ?? apporte un élément d'explication pour une situation musicale qui vient appuyer la perception de solidité du tempo : l'interprétation des musiciens renforce l'image d'une grille de pulsation – par analogie avec une fréquence fondamentale implicite – grâce à la tension interactive des placements rythmiques sur et autour de ces pulsations.

Afin d'avancer progressivement dans la complexité de l'élaboration d'un modèle, nous nous plaçons dans un premier temps dans le cadre d'un tempo relativement stable. Nous recherchons en (3.1) une grille de quantification de la séquence d'événements. La grille de quantification est une grille temporelle régulière d'intervalle ou *pas* Δ qui approche ou *quantifie* plus ou moins bien la séquence d'événements. L'objectif est ici de trouver la meilleure grille de quantification. Nous proposons en (3.1.2) une méthode de résolution géométrique de la grille que nous caractériserons par son *pas* Δ . En (3.2) nous introduisons un nouveau paramètre s centre de la grille qui permet son positionnement relatif dans le temps. Nous démontrons premièrement un théorème (Th.3.2.4) qui nous permet d'établir un algorithme rapide pour la recherche optimale de s et Δ . Nous abordons en (3.3) la résolution d'un problème de minimisation d'une fonction de coïncidence de la grille. C'est l'existence et la recherche de ce minimum qui donnent la dénomination de *grille optimale*. Un second théorème sur la *grille optimale* généralise par la suite le précédent en (3.3.5).

3.1 Grille de quantification

Le tempo et la grille de quantification sont intimement liés, même si nous ne sommes pas en mesure de justifier comment, ou plutôt qui vient avant qui pour le moment – nous le ferons plus loin. Nous recherchons le *tactus* d'une phrase rythmique : la plus petite pulsation qui permet de quantifier la phrase. Si cette plus petite pulsation est révélée, nous pouvons la voir comme un élément quantificateur de la phrase rythmique, car elle représente une unité de pulsation sur laquelle est construit le rythme. La question de l'unicité du *tactus* n'est pas évidente, car la perception peut-être multiple et changer dynamiquement. Il y a ici une ambiguïté qui résulte de la dualité entre la perception implicite et la transcription explicite dans la relation de la musique entendue avec la pensée musicale et rythmique qui la produit. Ces notions peuvent cependant partager les mêmes types de représentation à la base desquelles intervient le *tactus*.

3.1.1 *Tactus* ou *pas* de la grille de quantification

[Seppanen, 2001] propose une recherche du *tactus* à partir des interonsets¹ d’une phrase constituée par des impulsions rythmiques. Dans un cas idéal et mécanique, où le tempo ne bouge absolument pas, les interonsets sont des multiples entiers du *tactus* et nous pouvons considérer celui-ci comme le plus grand commun diviseur (PGCD) des interonsets. Pour un cas réel, on peut considérer que les interonsets contiennent une déviation aléatoire. [Seppanen, 2001] définit une fonction d’erreur restante, fonction d’une période Δ – notre *tactus* – et de la séquence d’interonsets o_i pour estimer le *tactus*.

$$e(\Delta) = \sum_{i=1}^n \left(\frac{o_i}{\Delta} - \left\lfloor \frac{o_i}{\Delta} + \frac{1}{2} \right\rfloor \right)^2 \quad (3.1.1)$$

Les minimums locaux de (3.1.1) représentent des candidats possibles pour le *tactus*. Si un PGCD exact existe, il peut être trouvé en cherchant la plus grande valeur Δ pour laquelle $e(\Delta) = 0$

$$pgcd(o_1, o_2, \dots, o_n) = \max\{\Delta | e(\Delta) = 0\} \quad (3.1.2)$$

Seppanen introduit un lissage des interonsets avec un histogramme des classes d’interonsets et fait une approximation de la valeur du meilleur *tactus* Δ^* cherché comme le plus grand minimum local trouvé sous un seuil déterminé par un coefficient choisit empiriquement que multiplie la valeur moyenne de $e(\Delta)$. Il n’est pas évident de décider du choix de ce seuil pour approximer Δ^* et la lisibilité du graphe $e(\Delta)$ ne permet pas toujours de lever les ambiguïtés.

La difficulté vient du fait que nous ne pouvons pas minimiser (3.1.1) par les méthodes classiques de calcul, car la partie entière qui apparait dans cette fonction la rend non différentiable : ses minimums ne sont pas accessibles avec une résolution analytique. Pour estimer le *tactus*, nous nous intéressons alors à une fonction comparable et aux propriétés géométriques liées à la présence de la partie entière qu’elle comporte. Nous désirons obtenir les minimums exacts et exhaustifs de la fonction.

¹un interonset est l’intervalle de temps entre deux “onsets” ou événements, la plupart du temps un *onset* représente l’instant de l’attaque d’une note

3.1.2 Résolution géométrique

Un travail de recherche mathématique en collaboration avec le professeur Bruno Simeone [Simeone, 2006] nous a permis d'orienter la résolution du problème d'approximation de la grille optimale de quantification vers une résolution géométrique et d'élaborer les algorithmes que nous décrivons par la suite.

Soit n points positifs sur l'axe des réels, support de nos événements temporels :

$$t_1 < t_2 < \dots < t_n$$

On désire trouver une grille régulière à une dimension qui donne la *meilleure approximation* de ces points.

Nous supposons que la grille contient l'origine et que son *pas* est Δ , i.e. la grille est formée par les points :

$$0, \Delta, 2\Delta, \dots, (p-1)\Delta, p\Delta$$

où p est le nombre de points de la grille pris en considération, en plus de l'origine (Fig.3.1). On désire choisir le pas Δ de façon à minimiser une fonction f qui représente la somme des distances des points donnés aux points de grille correspondants les plus proches. Formellement, on cherche un Δ^* qui minimise f tel que :

$$f(\Delta) = \sum_{k=1}^n \min_{h=0,1,\dots,p} |t_k - h\Delta| \quad (3.1.3)$$

Posons alors pour chaque point t_k la fonction f_k qui minimise sa distance à la grille :

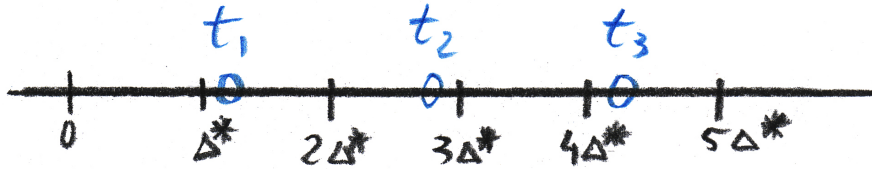


FIG. 3.1 – exemple de grille avec $k = 3$ et $p = 5$

$$f_k(\Delta) = \min_{h=0,1,\dots,p} |t_k - h\Delta| \quad (3.1.4)$$

Et introduisons la fonction principale :

$$g(\Delta) = \min_{h=0,1,\dots,p} |1 - h\Delta| \quad (3.1.5)$$

En comparant (3.1.4) et (3.1.5), on obtient l'identité :

$$f_k(\Delta) = t_k g\left(\frac{\Delta}{t_k}\right) \quad k = 1, \dots, n \quad (3.1.6)$$

Ainsi, la fonction f peut être ré-écrite de la manière suivante :

$$f(\Delta) = \sum_{k=1}^n t_k g\left(\frac{\Delta}{t_k}\right) \quad (3.1.7)$$

Une autre expression intéressante de f tient à l'hypothèse

$$\Delta \leq t_1 < t_2 < \dots < t_n < p\Delta \quad (3.1.8)$$

c'est à dire

$$\frac{t_n}{p} < \Delta \leq t_1 \quad (3.1.9)$$

Nous devons ici introduire quelques notations et définitions préliminaires.

Etant donnés deux nombres réels positifs t et d , il existe un entier unique q qui satisfait la condition

$$qd \leq t < (q+1)d \quad (3.1.10)$$

en fait, (3.1.10) est vérifié si et seulement si

$$q = \left\lfloor \frac{t}{d} \right\rfloor \quad (3.1.11)$$

où $\lfloor x \rfloor$ désigne la partie entière du réel x , i.e. x arrondi à l'entier inférieur.

Nous appelons q le quotient entier de t et d , et nous utilisons la notation :

$$q = \left\lfloor \frac{t}{d} \right\rfloor \equiv \text{quot}(t, d) \quad (3.1.12)$$

Notons alors que sous l'hypothèse (3.1.10) – ou son équivalent (3.1.11) – nous avons

comme illustré (Fig.3.2)

$$f_k(\Delta) = \min_{h=0,1,\dots,p} |t_k - h\Delta| = \min\{t_k - \text{quot}(t_k, \Delta)\Delta, (\text{quot}(t_k, \Delta) + 1)\Delta - t_k\} \quad (3.1.13)$$

On obtient alors l'expression :

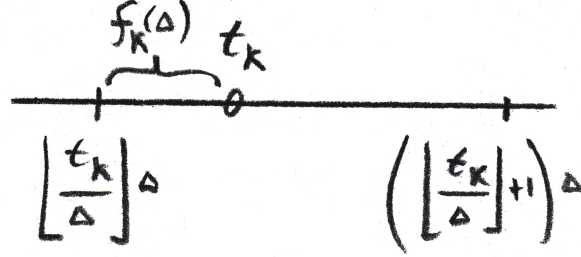


FIG. 3.2 – $f_k(\Delta)$

$$f(\Delta) = \sum_{k=1}^n \min \left\{ t_k - \left\lfloor \frac{t_k}{\Delta} \right\rfloor \Delta, \left(\left\lfloor \frac{t_k}{\Delta} \right\rfloor + 1 \right) \Delta - t_k \right\} \quad (3.1.14)$$

valide d'après (3.1.9) pour

$$\frac{t_n}{p} < \Delta \leq t_1$$

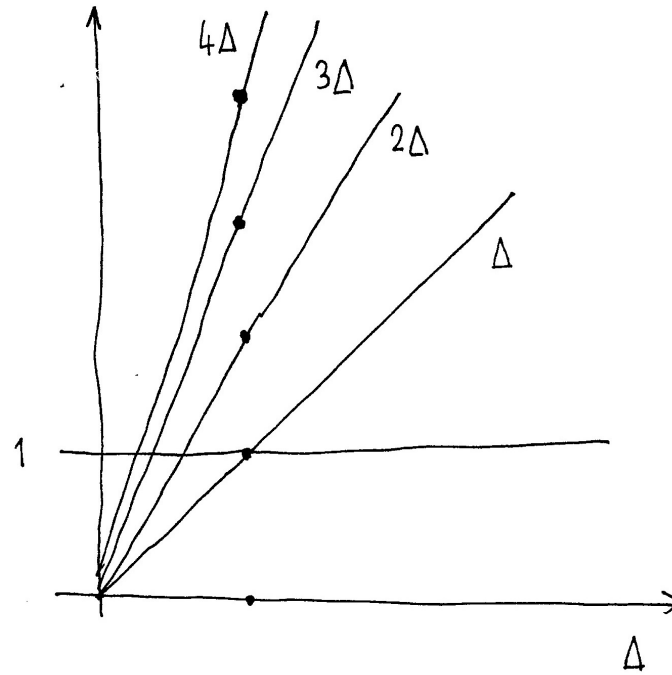
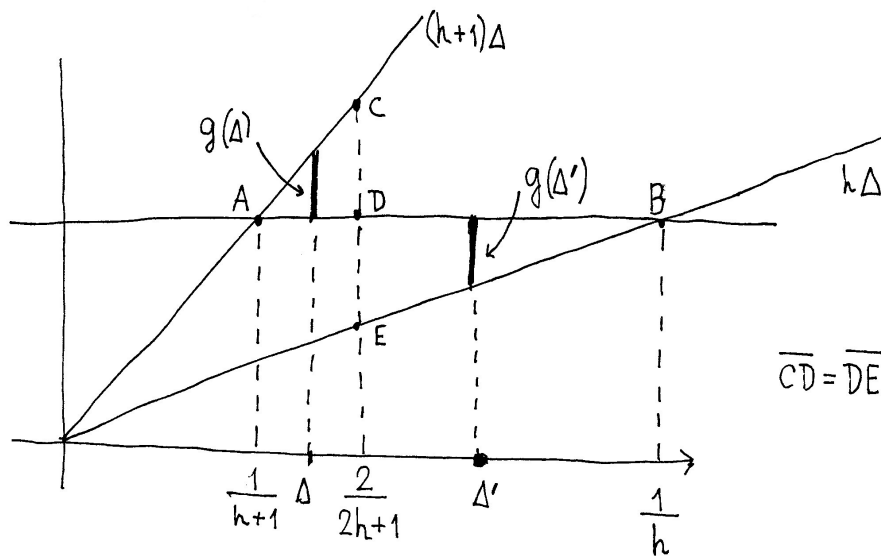
Etudions maintenant la fonction principale $g(\Delta)$ Fig.3.3.

D'après les figures Fig.3.3 et Fig.3.4, nous voyons que $g(\Delta)$ est une fonction linéaire par morceau.

$$g(\Delta) = \begin{cases} 1 - p\Delta & \text{si } 0 < \Delta \leq \frac{1}{p} \\ (h+1)\Delta - 1 & h = 1, \dots, p-1 \quad \text{si } \frac{1}{h+1} \leq \Delta \leq \frac{2}{2h+1}^* \\ 1 - h\Delta & h = 1, \dots, p-1 \quad \text{si } \frac{2}{2h+1} \leq \Delta \leq \frac{1}{h} \\ \Delta - 1 & \text{si } 1 \leq \Delta \end{cases} \quad (3.1.15)$$

Remarquons que $\frac{2}{2h+1}^*$ est la moyenne harmonique de $\frac{1}{h}$ et de $\frac{1}{h+1}$.

La figure Fig.3.5 illustre le graphe de $g(\Delta)$ pour $p = 4$.

FIG. 3.3 – $g(\Delta)$ FIG. 3.4 – zoom de $g(\Delta)$

Les points singuliers de g sont $\beta_1 > \beta_2 > \dots > \beta_{2p-1}$, où

$$\begin{cases} \beta_{2h-1} = \frac{1}{h} & h = 1, \dots, p \\ \beta_{2h} = \frac{2}{2h+1} & h = 1, \dots, p-1 \end{cases} \quad (3.1.16)$$

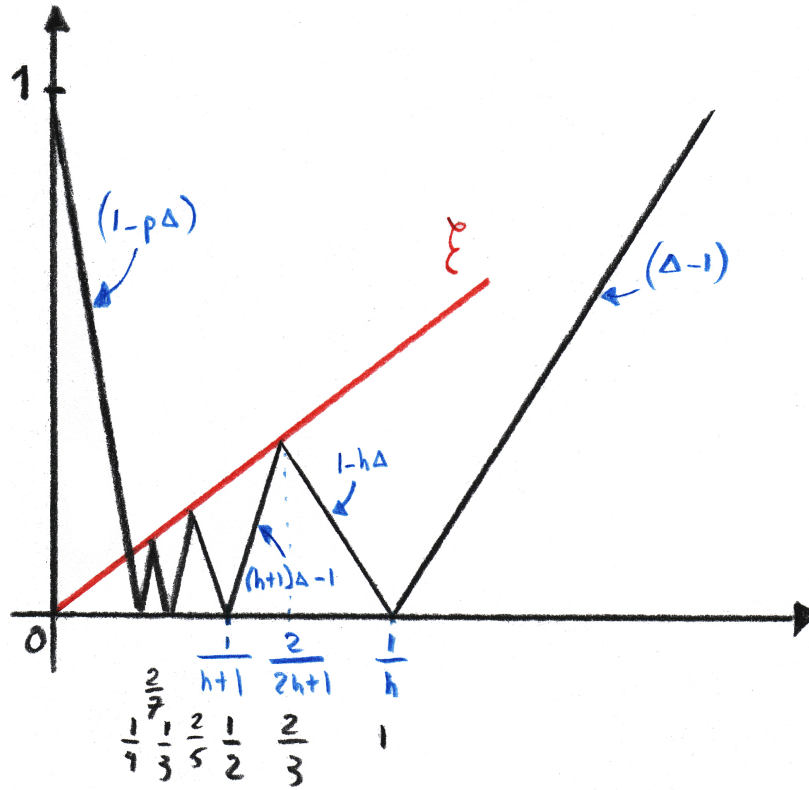


FIG. 3.5 – $g(\Delta)$ avec $p = 4$ (minimums pour $\frac{1}{h}$ en $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1$ et maximums pour $\frac{2}{2h+1}$ en $(\frac{2}{7}, \frac{2}{5}, \frac{2}{3})$)

On a alors

$$\begin{cases} g(\beta_{2h-1}) = 0 & h = 1, \dots, p \\ g(\beta_{2h}) = \frac{1}{2h+1} & h = 1, \dots, p-1 \end{cases} \quad (3.1.17)$$

Nous voyons alors que g admet en alternance parmi ses $2p-1$ points singuliers p minimums locaux et $p-1$ maximum locaux.

En comparant (3.1.16) et (3.1.17), on voit que les points de coordonnées $(\frac{2}{2h+1}, \frac{1}{2h+1})$ appartiennent au graphe de g et sont tous alignés sur la droite $\xi = \frac{1}{2}\Delta$.

Soit $(\Delta, g(\Delta))$ un point arbitraire du graph g . D'après (3.1.6), nous avons :

$$(t_k \Delta, f_k(t_k \Delta)) = (t_k \Delta, t_k g(\Delta)) = t_k(\Delta, g(\Delta)) \quad k = 1, \dots, n \quad (3.1.18)$$

C'est à dire que le mapping linéaire ou l'homothétie de rapport t_k et de centre $(0, 0)$:

$$\begin{cases} \Delta' = t_k \Delta \\ \xi' = t_k \xi \end{cases} \quad (3.1.19)$$

transforme le graph de g en celui de f_k pour chaque k . En particulier, les points singuliers de f_k sont les images de ceux de g sous (3.1.19). Ils sont donnés par :

$$\alpha_{kh} = t_k \beta_h \quad h = 1, \dots, 2p - 1 \quad (3.1.20)$$

et finalement , les points singuliers de f sont donnés par :

$$\alpha_{kh} = t_k \beta_h \quad k = 1, \dots, n \quad h = 1, \dots, 2p - 1 \quad (3.1.21)$$

car α est un point singulier de f si et seulement s'il est un point singulier d'un f_k .

Nous remarquons au passage que f admet en tout $n(2p-1)$ points singuliers. f est une fonction linéaire par morceaux car somme des fonctions f_k linéaires par morceaux. Sur l'intervalle constitué par les α_{kh} nous pouvons reconstruire f en une suite de segments de droite joignant les points $(\alpha_{kh}, f(\alpha_{kh}))$ après un réordonnement croissant des α_{kh} .

Pour notre propos, la signification des points singuliers de f tient dans le fait que, puisque f est linéaire par morceau, ses minimums sont atteints parmi ses points singuliers. Ceci est une conséquence directe du fait que toute fonction linéaire sur un intervalle fermé atteint son minimum – et son maximum – en l'une des extrémités de l'intervalle.

Néanmoins, afin d'obtenir une solution significative de $f(\Delta)$, nous devons restreindre les valeurs de Δ dans un intervalle convenable. En premier lieu, si nous choisissons Δ arbitrairement petit et p suffisamment grand, nous pouvons rendre la grille de points arbitrairement dense sur la ligne, et nous pouvons approximer les points donnés par les points de la grille avec la précision – ou la résolution – voulue. En pratique, cependant, nous poserons une borne inférieure c pour Δ :

$$\Delta \geq c \quad (3.1.22)$$

Considérant la pulsation, le *tactus* ne peut pas être trop petit. En ce cas, il est aussi

raisonnable d'imposer une borne supérieure à Δ afin que le *tactus* ne puisse pas devenir trop grand. En second lieu, nous pourrions désirer que le dernier point donné soit entouré par les deux derniers points de la grille :

$$(p-1)\Delta \leq t_n < p\Delta \quad (3.1.23)$$

Ou bien son équivalent

$$\frac{t_n}{p} < \Delta \leq \frac{t_n}{p-1} \quad (3.1.24)$$

Remarquons que si (3.1.22) et la seconde inégalité (3.1.24) sont réunies, on a

$$(p-1)c \leq t_n \quad (3.1.25)$$

et par conséquent

$$p \leq p_{max} \equiv 1 + \left\lfloor \frac{t_n}{c} \right\rfloor \quad (3.1.26)$$

De cette manière on obtient une borne supérieure sur le nombre de points de la grille. Quand p est grand, il se peut que l'intervalle (3.1.24) devienne très petit. Comme alternative, on peut imposer la contrainte (3.1.9). Notons qu'une condition nécessaire pour que l'intervalle (3.1.9) ne soit pas vide est

$$\frac{t_n}{p} < t_1 \quad \text{i.e.} \quad p > \frac{t_n}{t_1} \quad (3.1.27)$$

Si t_1 est aussi plus grand que $\frac{t_n}{p-1}$, alors la contrainte (3.1.9) est plus faible que (3.1.24). De toute manière, en imposant la borne inférieure (3.1.22), et possiblement (3.1.9) ou (3.1.24), on obtient un intervalle I de valeurs possibles pour Δ . Il s'en suit alors la procédure Alg.3.1 pour minimiser f sur I .

Algorithm 3.1 Minimisation de f sur I

données

$\{t_1, \dots, t_n\}$, séquence de n points

p , nombre de points de grille

I , Intervalle de faisabilité de Δ

générer les $n(2p-1)$ points singuliers de f d'après (3.1.16) et (3.1.21)

calculer la valeur de f pour chaque point singulier avec (3.1.14), en ayant éliminé ceux qui ne font pas partie de I

return le point singulier pour lequel la valeur de f est la plus petite est un point minimum de f sur I

La procédure **Alg.3.1** met en oeuvre un nombre maximum de $n^2(2p - 1)$ itérations suivant les points retenus.

3.1.3 Remarques

Nous avons montré que l'on peut décomposer la fonction de minimisation de distance f (3.1.3) en une somme de fonctions f_k (3.1.4) positives, toutes homothétiques une fonction principale g (3.1.5) positive et linéaire par morceaux. Nous avons montré aussi que pour un nombre de points p dont on choisit judicieusement la valeur, g admet un nombre fini de minimums et de maximums locaux qui donnent les minimums et les maximums locaux des f_k de par leurs relations homothétiques. Remarquons que les minimums locaux de f se trouvent parmi les minimums locaux de ses f_k à la condition que tous les α_{kh} soient distincts (on peut ne pas tenir compte des maximums locaux des f_k dans ce cas).

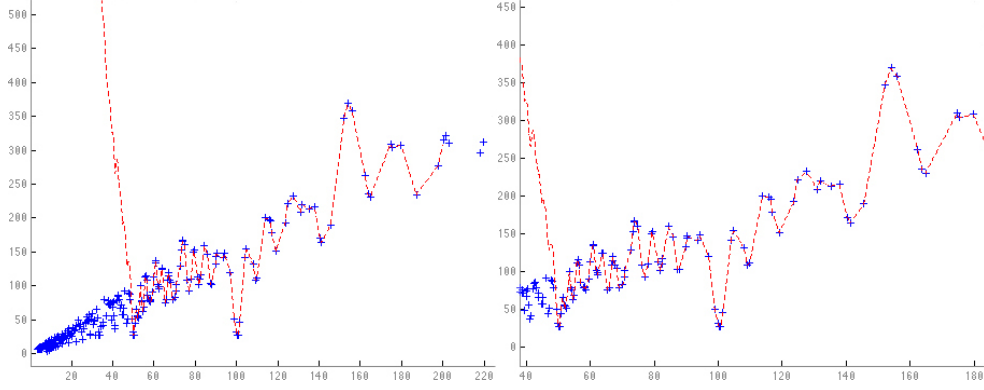
Le calcul des α_{kh} (3.1.20) permet de discrétiser le nombre de candidats possibles pour trouver le meilleur *pas* Δ^* , et de s'affranchir des approximations d'une résolution purement numérique comme celle proposée par [Seppanen, 2001]. Le nombre de points de grille p varie et engendre un nombre fini de candidats Δ^* pour minimiser f . En effet, on sait de manière intuitive que s'il existe une grille de quantification, seules certaines valeurs Δ du *pas* permettent de quantifier une phrase rythmique. Ainsi, il n'est pas nécessaire de faire varier continûment cette valeur : la double croche, la croche ou leurs équivalents ternaires sont souvent ce que le musicien prend pour référence quand il joue.

Notons également la propriété suivante : pour p donné, d'après (3.1.17) et (3.1.20), l'ensemble des $\{\alpha_{kh}\}_p$ points singuliers de f , est un sous ensemble de $\{\alpha_{kh}\}_{p+1}$ pour une grille à $p + 1$ points. Ce qui signifie qu'on conserve les candidats Δ^* potentiels quand on augmente p , i.e. la résolution de la grille. Il y a conservation des points singuliers d'une valeur de p à la suivante et on ne fait que rajouter des discontinuités au graphe de f en augmentant p

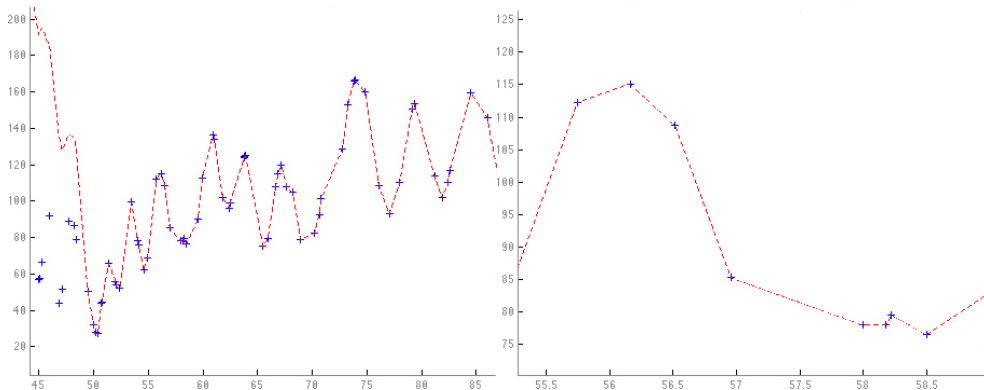
$$\{\alpha_{kh}\}_p \subset \{\alpha_{kh}\}_{p+1} \quad (3.1.28)$$

Les figures **Fig.3.6(a)** **Fig.3.7(a)** et **Fig.3.6(b)** mettent en oeuvre l'algorithme **Alg.3.1** pour une séquence de 6 points en entrée – (100, 203, 406, 702, 989, 1310). Le calcul des

points singuliers – les points (+) – reconstruisent le graphe de f (3.1.3) qu’un calcul direct – la courbe en trait discontinus – ne permet pas de donner quand on s’approche trop de l’origine, à moins d’une précision trop grande et prohibitive par son coup de calcul. Il est de toute façon impossible de trouver la valeur exacte des points singuliers par une simple approximation numérique de f (3.1.3).



(a) Reconstruction de f avec $p = 26$. les candidats pour Δ^* sont 50.4 et 100.8 auxquels on peut penser à la vue des valeurs de la séquence.



(b) Zoom consécutifs sur les points singuliers illustrant la reconstruction de f , en traits discontinus le calcul par approximation numérique et en croix + les α_{hk} .

FIG. 3.6 – Calcul des points singuliers pour la séquence de points (100, 203, 406, 702, 989, 1310)

La figure Fig.3.7 illustre la grille correspondant à cette même série de points. Le meilleur pas dans l’intervalle de recherche considéré – $[40, 1000]$ – est de 100.8 ms, on voit qu’un pas de 100 ms permet à deux notes – première et cinquième – d’être très proches d’un point de grille, alors qu’un pas de 101 ms permet à quatre notes – première, troisième, quatrième et dernière – de se rapprocher de leur points de grilles. Même une faible différence sur la valeur du pas de grille se propage sur toute la durée de la séquence d’entrée et entraîne une erreur visible accumulée.

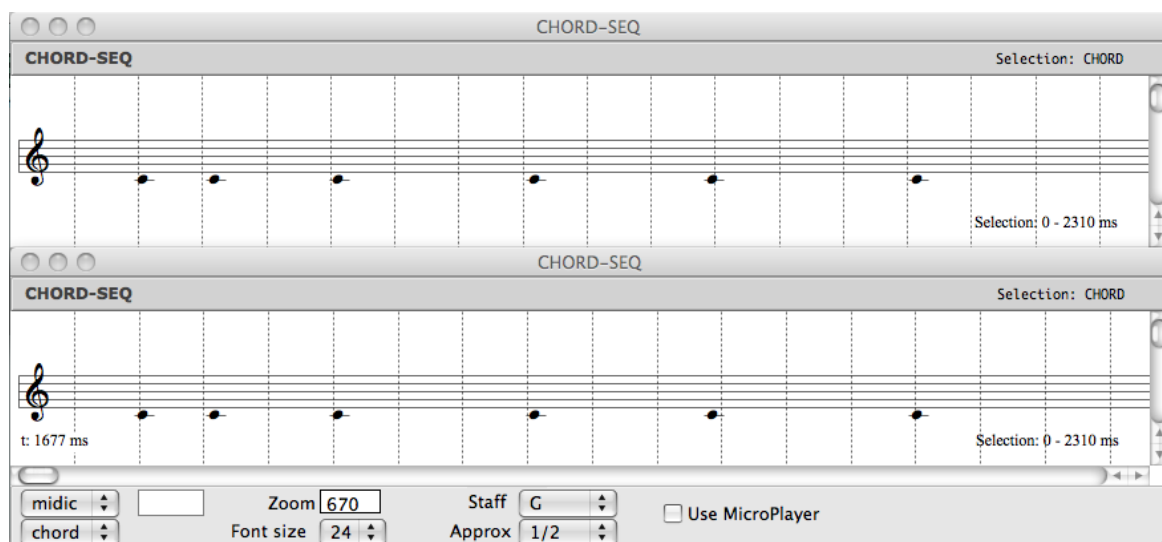


FIG. 3.7 – Grilles de 100 et 101 ms pour la séquence (100, 203, 406, 702, 989, 1310)

3.2 Meilleure phase pour une grille de quantification optimale

3.2.1 Reformulation du problème

Nous avons montré comment trouver la meilleure grille de quantification pour un ensemble de points positifs $\{t_n\}$. Cette grille est une suite de points régulièrement espacés, qui *collent* ou *approximent* du mieux possible les points de la suite $\{t_n\}$. L'idée est que la grille *coïncide* au mieux avec la suite de points donnés, ou complètement dans un cas idéal.

Pour le moment, nous avons posée comme hypothèse pour la résolution mathématique du problème que de la grille de quantification passe par l'origine car c'est une hypothèse que . Néanmoins, si elle est optimale dans ce cas, on ne peut affirmer qu'elle le soit si on désire qu'elle ne passe plus par l'origine. Nous constatons ainsi que cet état *optimal* est relatif car il ne dépend pas seulement de la variable Δ .

Autrement dit, à première vue, il peut paraître assez trivial que l'extension du problème pour lequel les points de la grille sont $\{s, s + \Delta, s + 2\Delta, \dots, s + p\Delta\}$, avec $s \in \mathbb{R}$ puisse être envisagée à partir de cette même démarche. En effet, pourquoi ne pas chercher la meilleure *phase* de la grille de *tactus* engendrée par le candidat Δ retenu ? Cela

pourrait être une bonne approximation de la position de la grille de quantification optimale en considérant, ou plutôt en espérant, s petit devant Δ . En fait, rien ne nous permet d'affirmer que si Δ est une valeur optimale pour la suite $\{t_n\}$, il l'est aussi pour la suite $\{t_n - s\}$. Pourquoi ? Parce-que notre résolution implique le point origine qui nous gêne maintenant.

Finalement, nous pouvons dire que la recherche de la meilleure phase d'une *grille de quantification optimale* nécessite une reformulation du problème : nous cherchons le meilleur couple $(phase, \Delta(phase))$ ou $(\Delta, phase(\Delta))$ car notre méthode ne permet pas directement d'optimiser un paramètre puis le second. Les variables s et Δ sont liées, et la grille optimale que nous cherchons est donnée par le couple optimal (s^*, Δ^*) .

Nous présentons ci-après une extension du problème de minimisation de la fonction f étudiée précédemment, mettant en oeuvre maintenant deux variables, s et Δ , et définie sur un domaine qui est une *surface fermée continue* que nous préciserons. Nous donnons par la suite une méthode qui permet de discrétiser ce domaine en un ensemble dénombrable et fini de points. Nous en déduisons un nouvel algorithme pour la résolution numérique optimale exacte et non approchée du couple (s, Δ) dans son domaine de définition. Nous ne désirons pas une exactitude absolue, même si nous l'obtenons, nous voulons seulement ne pas passer à côté du résultat optimal, ce que nous ne pouvons espérer avec une méthode numérique approchée car les équations en jeu ne sont pas dérivables.

3.2.2 Périodicité et phase optimales

Nous présentons ici les éléments de démonstration mathématique qui nous permettent de poursuivre la recherche de la *grille optimale de quantification*. Nous devons montrer premièrement l'existence de la périodicité de la fonction à minimiser, et de la phase optimale qui la minimise.

Nous nous proposons de trouver le centre d'une grille de taille de pas Δ donné, afin qu'elle coïncide au mieux avec une suite de points $\{t_n\}$.

Soit $0 < t_1 \leq t_2 \leq \dots \leq t_n$ n points sur l'axe des réels. Considérons la grille unidimensionnelle de centre s et de taille de pas $\Delta > 0$:

$$\dots, s - 2\Delta, s - \Delta, s, s + \Delta, s + 2\Delta, \dots$$

Etant donné un pas de grille Δ , nous voulons trouver un centre s^* qui minimise la fonction de phase φ de la grille.

$$\varphi(s) = \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - h\Delta| \quad (3.2.1)$$

où $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ est l'ensemble des entiers.

Lemme 3.2.1. *La fonction $\varphi(s)$ est périodique de période Δ*

$$\varphi(s + \Delta) = \varphi(s), \forall s \quad (3.2.2)$$

Démonstration :

$$\begin{aligned} \varphi(s + \Delta) &= \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - \Delta - h\Delta| \\ &= \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - (h + 1)\Delta| \\ &= \sum_{k=1}^n \min_{h' \in \mathbb{Z}} |t_k - s - \Delta - h'\Delta| \\ &= \varphi(s) \end{aligned}$$

□

En regard du Lemme 3.2.1, il est suffisant de calculer le minimum de $\varphi(s)$ pour $0 \leq s < \Delta$

Posons

$$M_k = \min \{h \in \mathbb{Z} : t_k - h\Delta \leq 0\} \quad (3.2.3)$$

$$m_k = \max \{h \in \mathbb{Z} : t_k - h\Delta \geq \Delta\} \quad (3.2.4)$$

Remarquons que $M_k - m_k \leq 2$

alors, pour $0 \leq s < \Delta$, nous avons

$$\varphi(s) = \sum_{k=1}^n \min_{m_k \leq h \leq M_k} |t_k - s - h\Delta| \quad (3.2.5)$$

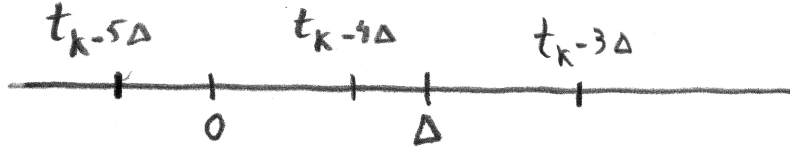


FIG. 3.8 – exemple avec $m_k = 3$ et $M_k = 5$ pour le calcul de $\varphi(s)$

Lemme 3.2.2. *La fonction $\varphi(s)$ est linéaire par morceaux dans l'intervalle $0 \leq s < \Delta$ et a toujours un point minimum de la forme*

$$s^* = t_q - l\Delta \quad \text{avec} \quad q = 1, \dots, n \quad \text{et} \quad l = m_q, \dots, M_q$$

Démonstration :

Pour h et k fixés, la fonction

$$\lambda_{hk}(s) = |t_k - s - h\Delta| \quad (3.2.6)$$

s'annule en $t_k - h\Delta$ (Fig. 3.9)

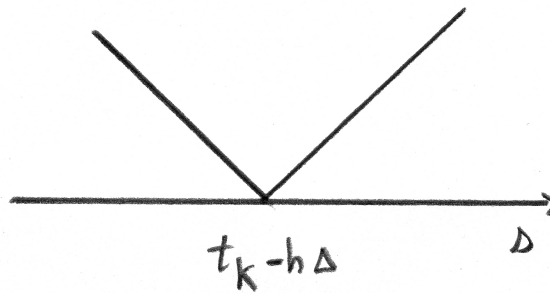
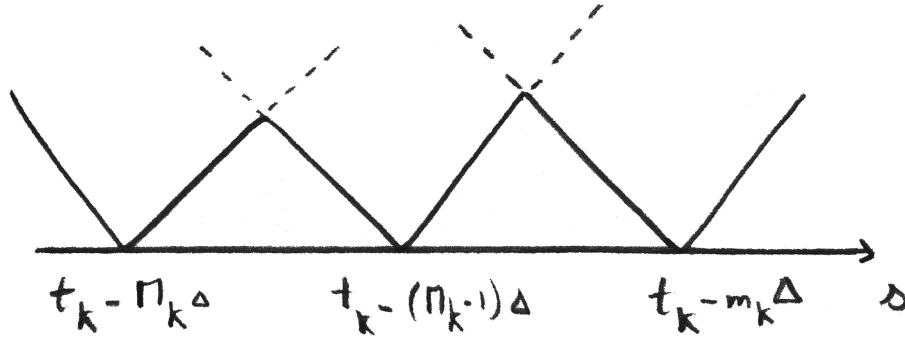


FIG. 3.9 – $\lambda_{hk}(s)$

et pour k fixé, posons (Fig. 3.10)

$$\mu_k(s) = \min_{m_k \leq h \leq M_k} |t_k - s - h\Delta| \quad (3.2.7)$$

FIG. 3.10 – $\mu_k(s)$

Finalement, $\varphi(s)$ est la somme de n fonctions de la forme (3.2.7) et donc est linéaire par morceau et atteint son minimum en un de ses points anguleux $t_q - l\Delta$, $m_q \leq l \leq M_q$.

□

Nous nous rappelons que, étant donné t , $d > 0$, il existe q et r uniques, où $q \in \mathbb{Z}$ et $0 \leq r < d$ tels que $t = qd + r$. q est le quotient entier $quot(t, d)$ et r la partie fractionnaire $frac(t, d)$ Fig.??.

$$q = \left\lfloor \frac{t}{d} \right\rfloor, \quad r = t - \left\lfloor \frac{t}{d} \right\rfloor d \quad (3.2.8)$$

Théorème 3.2.3. *La fonction $\varphi(s)$ atteint son minimum en l'un des n points*

$$frac(t_k, \Delta), \quad k = 1, \dots, n$$

Démonstration :

D'après le Lemme 3.2.2, le minimum de $\varphi(s)$ est atteint en un point de la forme $s^* = t_q - l\Delta$. Comme ce point doit appartenir à l'intervalle $0 \leq s < \Delta$, on a

$$0 \leq t_q - l\Delta < \Delta$$

$$l \leq \frac{t_q}{\Delta} < l + 1$$

d'où

$$l = \left\lfloor \frac{t_q}{\Delta} \right\rfloor \quad \text{et} \quad t_q - l\Delta = \text{frac}(t_q, \Delta) = s^*$$

□

Théorème 3.2.4. *Soit une grille de taille de pas donné Δ et de centre s^* , qui coïncide au mieux avec une suite de points $\{t_n\}$, alors il existe au moins un entier q^* , $1 \leq q^* \leq n$, tel que t_{q^*} soit confondu avec un point de la grille.*

Démonstration :

D'après le théorème 3.2.3, si la grille coïncide au mieux avec $\{t_n\}$, alors s^* est un des n éléments de l'ensemble des $\text{frac}(t_k, \Delta)$, $1 \leq k \leq n$, i.e.

$$\exists \quad q^* \quad , \quad 1 \leq q^* \leq n \quad : \quad s^* = \text{frac}(t_{q^*}, \Delta)$$

la distance du point t_{q^*} au point de grille le plus proche est donnée par

$$\begin{aligned} & \min_{h \in \mathbb{Z}} |t_{q^*} - s^* - h\Delta| \\ &= \min_{h \in \mathbb{Z}} |t_{q^*} - \text{frac}(t_{q^*}, \Delta) - h\Delta| \\ &= \min_{h \in \mathbb{Z}} |l^*\Delta - h\Delta| \\ &= 0 \end{aligned} \quad \text{en prenant } h = l^*$$

i.e. t_{q^*} est confondu avec un point de la grille et $s^* = t_{q^*} - l^*\Delta$.

□

Nous discuterons par la suite des implications de ce théorème.

3.2.3 Algorithme de minimisation de la phase de la grille

Nous pouvons maintenant déduire un algorithme Alg.3.2 de calcul du minimum de $\varphi(s)$ pour un Δ donné. La complexité est en $\mathcal{O}(n^2)$.

Algorithm 3.2 minimum de $\varphi(s)$ pour un Δ donné

données $\{t_1, \dots, t_n\}$, séquence de n points Δ , pas de la grille**for** $q = 1$ to n **do** $s_q \leftarrow \text{frac}(t_q, \Delta)$, centre candidat possible $\phi_q \leftarrow \sum_{k=1}^n \min \{ \text{frac}(t_k - s_q, \Delta), \Delta - \text{frac}(t_k - s_q, \Delta) \}$ avec $\text{frac}(x, y) = x - \left\lfloor \frac{x}{y} \right\rfloor y$ **end for** q^* tel que $\phi_{q^*} \leftarrow \min_{q=1, \dots, n} \{ \phi_q \}$ **return** ϕ_{q^*} , qui est le $\min_{0 \leq s < \Delta} \phi(s) = \phi(s^*)$ i.e. le centre optimalet t_{q^*} est confondu avec un point de la grille d'après le théorème [Th.3.2.4](#).

3.2.4 Domaine de définition de la fonction à optimiser

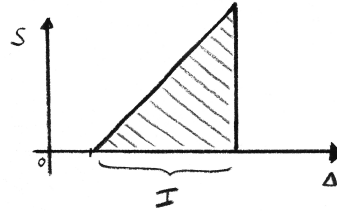
Nous avons étendu le problème unidimensionnel (3.1.3) à (3.2.1) en montrant l'existence d'un centre s^* qui engendre la phase minimale de la grille. Nous devons maintenant minimiser la fonction de deux variables $\varphi(s, \Delta)$

$$\varphi(s, \Delta) = \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - h\Delta| \quad , \quad \Delta \in I \quad (3.2.9)$$

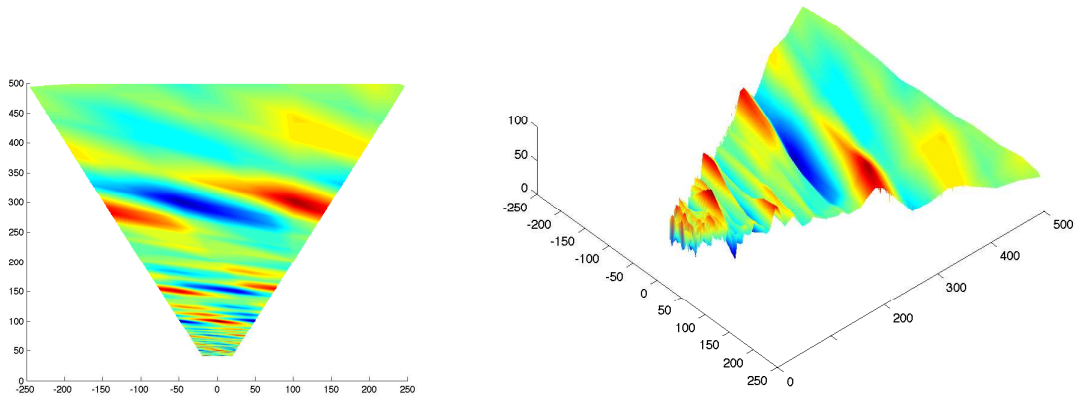
le Lemme 3.2.1 nous a permis d'affirmer qu'il est suffisant de calculer le minimum de $\varphi(s, \Delta)$ pour $0 \leq s < \Delta$.

Soit I l'intervalle des valeurs possibles pour Δ , borné par une valeur proche du plus petit intervalle entre deux valeurs successives de la suite $\{t_n\}$. Pour $\Delta \in I$, le domaine de définition bidimensionnel sur lequel nous optimisons $\varphi(s, \Delta)$ est une surface triangulaire tronquée représentée [Fig.3.11](#).

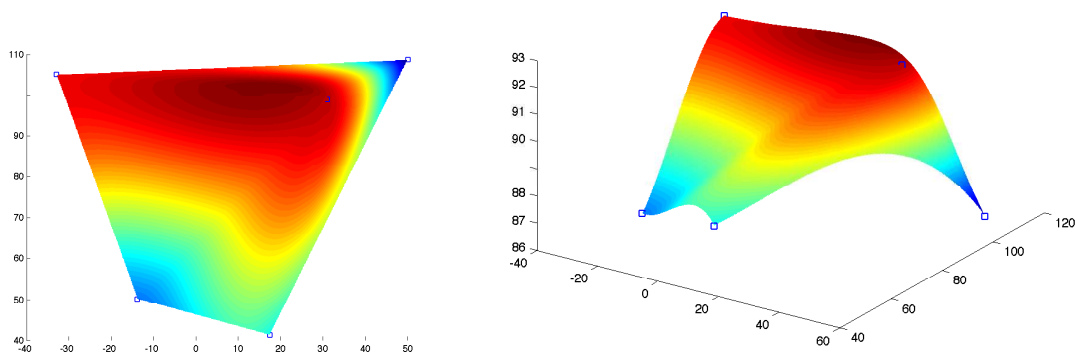
On sait minimiser $f(\Delta) = \varphi(0, \Delta)$, i.e. trouver les grilles unidimensionnelles passant par l'origine et engendrées par les candidats Δ possibles se trouvant parmi un ensemble fini dénombrable de $n(2p-1)$ valeurs. Nous nous retrouvons face à un problème d'optimisation bidimensionnelle sur une surface continue. Nous donnons la résolution de ce problème plus loin.

FIG. 3.11 – Surface domaine de définition pour minimiser $\varphi(s, \Delta)$.

Les figures Fig.3.12(a) et Fig.3.12(b) montrent une résolution numérique approchée de $\varphi(s, \Delta)$.



(a) Périodicité de f sur $\mathcal{D}(s, \Delta)$ et *indice de coïncidence*, le comportement devient chaotique quand on s'approche de l'origine.



(b) Détail pour l'optimisation du couple (s, Δ) avec *indice de coïncidence*.

FIG. 3.12 – Approximation de f sur $\mathcal{D}(s, \Delta)$ pour la séquence [100 203 406 702 989 1315].

3.2.5 Coïncidence optimale

Nous emploierons par la suite *coïncidence optimale* ou *meilleure approximation* de manière équivalente pour définir la notion suivante :

Définition 3.2.1. *Grille de coïncidence optimale*

Soit $\{t_n\}$ une suite de n points $0 \leq t_1 < t_2 < \dots < t_n$.

Soit $\Delta \in I$, intervalle donné, et $\mathcal{D}$ la surface triangulaire tronquée du plan (O, Δ, s) définie par la paramétrisation $\{ \Delta \in I, s \in [0, \Delta] \}$

Soit $\mathcal{G}_{s,\Delta}$ la grille unidimensionnelle de centre $s \in \mathbb{R}$ et de pas $\Delta > 0$, i.e. $\mathcal{G}_{s,\Delta} \equiv$ l'ensemble des points $\{s + m\Delta\}$, $m \in \mathbb{Z}$.

Soit $(s^*, \Delta^*) \in \mathcal{D}$

Nous appelons $\mathcal{G}_{s^*, \Delta^*}$ grille de coïncidence optimale de $\{t_n\}$ sur $\mathcal{D}$ ssi (s^*, Δ^*) minimise la fonction f telle que

$$f(s, \Delta) = \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - h\Delta| \quad , \quad (s, \Delta) \in \mathcal{D} \quad (3.2.10)$$

3.2.6 Séquence temporelle et repliement

Précédemment nous avons envisagé le calcul de la fonction de distance de la grille aux points de la séquence $\{t_n\}$ en pensant la translation s appliquée à la séquence et non pas au centre de la grille. La translation pensée pour la séquence implique une solution de grille centrée sur l'origine et de ce fait les deux variables s et Δ ne sont pas libres. La translation appliquée au centre de la grille implique une solution centrée sur la nouvelle origine donnée et rend les deux variables s et Δ indépendantes. Ce faisant, cela peut se voir comme un *repliement* de la séquence autour du nouveau point origine : nous aurions un calcul de 2 grilles optimales, réunies dans un calcul unique avec un réarrangement des points de la séquence **Fig.??**

Nous avons montré avec le **Th.3.2.4** que l'on peut discrétiser le calcul des solutions de cette nouvelle fonction s'appuyant sur la combinaison de deux grilles optimales dans l'espace continu de $(s, \Delta) \in \mathcal{D}$. La nouvelle grille optimale se trouve parmi les grilles centrées sur les points de la nouvelle séquence.

Nous observons que le *repliement* réalise un retournement progressif de la séquence : partant de la séquence d'origine quand le point de *repliement* est t_1 et arrivant à

la séquence rétrograde quand le point de *repliement* est t_n . Pouvons-nous mettre en relation cette façon de voir avec la perception consciente d'une séquence temporelle dans son déroulement sur l'axe du temps passé-présent, et la perception inconsciente de la même séquence rétrograde, i.e. dans son déroulement sur l'axe du temps inversé présent-passé ?

Nous parlons de perception inconsciente car il n'est pas facile de rejouer à l'envers un rythme perçu, alors qu'une symétrie visuelle est évidente. Néanmoins, si l'on considère la mémoire du musicien ou de l'auditeur, les événements les plus récents sont ceux reçus les derniers. Le rythme rétrograde devrait être paradoxalement plus évident dans sa forme mémorisée que le rythme direct, mais certainement c'est le poids de la flèche du temps qui l'emporte dans l'organisation de notre perception temporelle et notre mémoire consciente après traitement de cette information considère seulement le rythme direct. Nous manquons d'éléments de recherche psycho-cognitive dans la perception des intervalles de temps rétrograde pour pouvoir réellement répondre à cette hypothèse. Nous montrons plus loin en (3.3.1) comment nous utilisons le Th.3.2.4 pour rechercher la grille optimale sans utiliser la méthode de repliement évoquée ci-dessus qui aurait pour artefact de rajouter des points – et d'augmenter les calculs – qui ne sont pas nécessaires dans la séquence de points d'entrée.

3.3 Résolution bidimensionnelle

Un des résultats les plus significatifs que nous ayons obtenu, le Théorème Th.3.2.4, affirme qu'une grille de *coïncidence optimale* définie par la définition 3.2.1 contient toujours un des points de la séquence d'entrée. Sur la base de ce résultat, nous développons une procédure de complexité $\mathcal{O}(n^3 p^2)$ pour résoudre le problème bidimensionnel de coïncidence optimale que nous rappelons ci-dessous.

$$\min_{s, \Delta} f_p(s, \Delta) \quad \Delta \in I, \quad 0 \leq s \leq \Delta \quad (3.3.1)$$

$$\text{avec } f_p(s, \Delta) = \sum_{k=1}^n \min_{h=0, \dots, p} |t_k - (s + h\Delta)| \quad (3.3.2)$$

l'équation (3.3.2) décrit une grille de centre s et de pas Δ qui approxime les points t_k avec l'erreur de distance $f_p(s, \Delta)$. Par commodité, nous appelons *p-grille* une grille unidimensionnelle régulière qui compte p points en plus de son centre.

Soit (s^*, Δ^*) la solution optimale de (3.3.1), alors

$$s^*, s^* + \Delta^*, \dots, s^* + p\Delta^* \quad (3.3.3)$$

est une p -grille de coïncidence optimale.

D'après le Théorème Th.3.2.4, nous savons qu'il existe deux indexes $q^* \in \{1, \dots, n\}$ et $l^* \in \{0, \dots, p\}$ tels que

$$t_{q^*} = s^* + l^* \Delta^* \quad \text{avec} \quad l^* \in [0..p] \quad \text{et} \quad q^* \in [1..n] \quad (3.3.4)$$

q^* et l^* sont inconnus.

3.3.1 Paramétrisation de la grille optimale

L'idée est de paramétrer sur q et l , pour les choisir parmi toutes les possibilités. Exactement l points de la p -grille optimale sont à la gauche de t_q et exactement $p - l$ sont à la droite de t_q . Finalement, pour p fixé et pour un choix de q et l on peut trouver une grille de meilleure coïncidence de centre t_q , qui a l points à gauche de t_q et $p - l$ points à droite de t_q comme le montre la Fig.3.13 qui considère une séquence de six points.

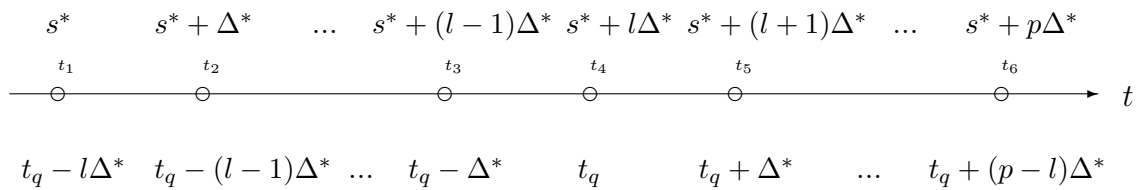


FIG. 3.13 – p -grille paramétrée sur q et l , la série d'événements est constituée par six points $\{t_1, t_2, t_3, t_4, t_5, t_6\}$ sur cet exemple. On considère ici $q = 4$ comme point confondu avec un point de grille

.

La taille de pas d'une telle grille provient de la résolution du problème de minimisation

unidimensionnel :

$$E_p(q, l) = \min_{\Delta \in I} G_p(\Delta; q, l) \quad (3.3.5)$$

$$\text{avec } G_p(\Delta; q, l) = \sum_{k=1}^n \min_{r=-l, \dots, p-l} |t_k - (t_q + r\Delta)| \quad (3.3.6)$$

L'interprétation de (3.3.6) est la suivante : la “fonction d'erreur” $G_p(\Delta; q, l)$ mesure la somme des distances des points de la séquence d'entrée $\{t_n\}$ aux points les plus proches de la grille de pas Δ définie ci-dessus.

Une fois le problème de minimisation (3.3.5) résolu pour tous les choix possibles de q et de l , on sélectionne le couple $(\bar{q}, \bar{l})$ pour lequel $E_p(\bar{q}, \bar{l})$ est minimal.

Soit $\bar{\Delta}$ le pas de grille correspondant, i.e.

$$E_p(\bar{q}, \bar{l}) = G_p(\bar{\Delta}; \bar{q}, \bar{l}) \quad (3.3.7)$$

et

$$\bar{s} = t_{\bar{q}} - \bar{l}\bar{\Delta} \quad (3.3.8)$$

Nous allons montrer que $(\bar{s}, \bar{\Delta})$ est une solution optimale de (3.3.1) , i.e.

$$f_p(\bar{s}, \bar{\Delta}) = f_p(s^*, \Delta^*) = \min_{(s, \Delta)} f_p(s, \Delta) \quad (3.3.9)$$

3.3.2 Algorithme d'optimisation bidimensionnelle

L'algorithme Alg.3.3 présente une description formelle de la procédure d'ensemble pour l'optimisation bidimensionnelle.

La complexité est en $\mathcal{O}(n(p+1))$ que multiplie la complexité de résolution de (3.3.5). Nous présenterons plus loin la procédure de résolution de cette équation. Pour le moment, nous allons démontrer l'exactitude de l'algorithme Alg.3.3.

Lemme 3.3.1. *Pour chaque $\Delta \in I$, on a*

$$G_p(\Delta; q, l) = f_p(t_q - l\Delta, \Delta) \quad (3.3.10)$$

Démonstration :

Algorithm 3.3 Meilleure grille d'approximation bidimensionnelle**données** $\{t_1, \dots, t_n\}$, séquence de n points p , nombre de points de grille en plus du centre I , intervalle de faisabilité du pas de grille**for** $q = 1$ to n **do****for** $l = 0$ to p **do**

, résolution du problème de minimisation de (3.3.5)

 $\Delta_{ql} \leftarrow$ pas de grille qui rend $E_p(q, l)$ minimum**end for****end for** $\bar{q}$ et $\bar{l}$ tel que $E_p(\bar{q}, \bar{l}) \leftarrow \min_{q=1, \dots, n; l=0, \dots, p} \{E_p(q, l)\}$ $\bar{\Delta} \leftarrow \Delta_{\bar{q}\bar{l}}$ $\bar{s} \leftarrow t_{\bar{q}} - \bar{l}\bar{\Delta}$ **return** $\bar{s}$ et $\bar{\Delta}$, centre et pas de la grille de coïncidence optimale

On a

$$\begin{aligned}
G_p(\Delta; q, l) &= \sum_{k=1}^n \min_{r=-l, \dots, p-l} |t_k - (t_q + r\Delta)| \\
&= \sum_{k=1}^n \min_{h=0, \dots, p} |t_k - (t_q + (h-l)\Delta)| \\
&= \sum_{k=1}^n \min_{h=0, \dots, p} |t_k - (t_q - l\Delta) - h\Delta| \\
&= f_p(t_q - l\Delta, \Delta) \quad \text{avec } f \text{ définie en (3.2.10)}
\end{aligned}$$

□

Théorème 3.3.2. Le couple $(\bar{s}, \bar{\Delta})$ résultat de l'algorithme Alg. 3.3 minimise $f_p(s, \Delta)$ (3.3.2)*Démonstration :*Supposons que (s^*, Δ^*) minimise $f_p(s, \Delta)$. Montrons que (3.3.9) est vrai.Clairement , $f_p(s^*, \Delta^*) \leq f_p(\bar{s}, \bar{\Delta})$ puisque (s^*, Δ^*) minimise $f_p(s, \Delta)$.D'autre part, d'après le Théorème 3.2.4, il existe q^* et l^* tels que $t_{q^*} = s^* + l^*\Delta^*$,

d'où

$$\begin{aligned}
f_p(\bar{s}, \bar{\Delta}) &= G_p(\bar{\Delta}; \bar{q}, \bar{l}) && \text{d'après le Lemme 3.3.1} \\
&= E_p(\bar{q}, \bar{l}) && \text{par définition de } \bar{\Delta} \\
&\leq E_p(q^*, l^*) && \text{puisque } (\bar{q}, \bar{l}) \text{ minimise } E_p(q, l) \\
&= G_p(\Delta_{q^*l^*}; q^*, l^*) && \text{par définition} \\
&\leq G_p(\Delta^*; q^*, l^*) && \text{puisque } \Delta_{q^*l^*} \text{ minimise } G_p(\Delta; q^*, l^*) \\
&= f_p(s^*, \Delta^*) && \text{d'après (3.3.10)}
\end{aligned}$$

d'où (3.3.9) est vrai.

□

Corollaire 3.3.3. *Considérons une p -grille de coïncidence optimale centrée sur s et un de ses points s' . Considérons une p -grille de coïncidence optimale centrée sur s' . Alors les pas des deux grilles coïncident.*

Démonstration :

Cela découle directement du Lemme 3.3.1

Corollaire 3.3.4. *Soit pour une séquence $\{t_n\}_{n=1,\dots,n}$ la p -grille de coïncidence optimale de centre s^* , de pas Δ^* , notée $\mathcal{G}_{s^*, \Delta^*}$. Soit la séquence $\{t'_n\}_{n=1,\dots,n}$ et $c \in \mathbb{R}$ telle que $\{t'_k\} = \{t_k - c\}_{k=1,\dots,n}$. Alors $\mathcal{G}_{s^*-c, \Delta^*} \equiv \mathcal{G}_{s'^*, \Delta'^*}$, i.e. la grille de coïncidence optimale d'une séquence translatée est la translation de la grille de coïncidence optimale de la séquence d'origine.*

Démonstration :

$\mathcal{G}_{s^*, \Delta^*}$ implique que (s^*, Δ^*) minimise

$$f_p(s, \Delta)_{(t)} = \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - s - h\Delta| \quad (3.3.11)$$

pour $\{t'_k\} = \{t_k - c\}_{k=1,\dots,n}$, sa p -grille optimale $\mathcal{G}_{s'^*, \Delta'^*}$ est telle que (s'^*, Δ'^*) minimise

$$\begin{aligned} f_p(s', \Delta')_{(t')} &= \sum_{k=1}^n \min_{h=1,\dots,p} |t'_k - s' - h\Delta'| \\ &= \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - c - s' - h\Delta'| \\ &= \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - (s' + c) - h\Delta'| \end{aligned}$$

soit en posant $s'' = s' + c$, (s'^*, Δ'^*) minimise

$$f_p(s', \Delta')_{(t')} = \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - s'' - h\Delta'| \quad (3.3.12)$$

Les membres droits de (3.3.11) et (3.3.12) étant identiques, le couple (s^*, Δ^*) qui minimise (3.3.11) minimise aussi (3.3.12) et donc

$$\begin{aligned} (s^*, \Delta^*) &\equiv (s''^*, \Delta'^*) \\ (s^*, \Delta^*) &\equiv (s'^* + c, \Delta'^*) \\ s^* &\equiv s'^* + c \text{ et } \Delta^* = \Delta'^* \end{aligned} \quad (3.3.13)$$

De (3.3.13) nous déduisons que $(s^* - c, \Delta^*)$ minimise (3.3.12). $\square$

Corollaire 3.3.5. *Soit pour une séquence $\{t_n\}_{n=1,\dots,n}$ la p -grille de coïncidence optimale de centre s^* , de pas Δ^* , notée $\mathcal{G}_{s^*, \Delta^*}$. Soit la séquence $\{t'_n\}_{n=1,\dots,n}$ et $\gamma \in \mathbb{R}^*$ telle que $\{t'_k\} = \{\gamma t_k\}_{k=1,\dots,n}$. Alors $\mathcal{G}_{\gamma s^*, \gamma \Delta^*}^* \equiv \mathcal{G}_{s'^*, \Delta'^*}$.*

Démonstration :

$\mathcal{G}_{s^*, \Delta^*}$ implique que (s^*, Δ^*) minimise

$$f_p(s, \Delta)_{(t)} = \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - s - h\Delta| \quad (3.3.14)$$

pour $\{t'_k\} = \{\gamma t_k\}_{k=1,\dots,n}$, sa p-grille optimale $\mathcal{G}_{s'^*, \Delta'^*}$ est telle que (s'^*, Δ'^*) minimise

$$\begin{aligned}
 f_p(s', \Delta')_{(t')} &= \sum_{k=1}^n \min_{h=1,\dots,p} |t'_k - s' - h\Delta'| \\
 &= \sum_{k=1}^n \min_{h=1,\dots,p} |\gamma t_k - s' - h\Delta'| \\
 &= \sum_{k=1}^n \min_{h=1,\dots,p} \left| \gamma \left(t_k - \frac{s'}{\gamma} - h \frac{\Delta'}{\gamma} \right) \right| \\
 &= \gamma \sum_{k=1}^n \min_{h=1,\dots,p} \left| t_k - \frac{s'}{\gamma} - h \frac{\Delta'}{\gamma} \right|
 \end{aligned} \tag{3.3.15}$$

soit en posant $s'' = \frac{s'}{\gamma}$, et $\Delta'' = \frac{\Delta'}{\gamma}$, (s'^*, Δ'^*) minimise

$$f_p(s', \Delta')_{(t')} = \sum_{k=1}^n \min_{h=1,\dots,p} |t_k - s'' - h\Delta''| \tag{3.3.16}$$

Les membres droits de (3.3.14) et (3.3.16) étant identiques, le couple (s^*, Δ^*) qui minimise (3.3.14) minimise aussi le membre droit de (3.3.16), or (s''^*, Δ''^*) minimise le membre droit de (3.3.16) :

$$\begin{aligned}
 (s^*, \Delta^*) &\equiv (s''^*, \Delta''^*) \\
 (s^*, \Delta^*) &\equiv \left(\frac{s'^*}{\gamma}, \frac{\Delta'^*}{\gamma} \right) \\
 s'^* &\equiv \gamma s^* \text{ et } \Delta'^* = \gamma \Delta^*
 \end{aligned} \tag{3.3.17}$$

De (3.3.17) nous déduisons que $(\gamma s^*, \gamma \Delta^*)$ minimise (3.3.14). $\square$

Remarque 1 : L'algorithme Alg.3.3 ne garantit pas que le centre $\bar{s}$ résultant appartienne à l'intervalle $[0, \Delta]$. Mais on peut facilement résoudre ce problème. Soit $\bar{\Delta}$ le pas de grille optimal et t_q le point d'entrée de la séquence $\{t_n\}$ appartenant à la grille de coïncidence optimale. D'après le Corrolaire 3.3.3, si s' est un point d'une telle grille, alors le couple $(s', \bar{\Delta})$ minimise $f_p(s, \Delta)$. Nous choisissons alors :

$$s' = t_q - \left\lfloor \frac{t_q}{\bar{\Delta}} \right\rfloor \bar{\Delta} \quad (3.3.18)$$

$$= \text{frac}(t_q, \bar{\Delta}) \quad (3.3.19)$$

Le point s' appartient à l'intervalle $[0, \Delta]$.

Remarque 2 : Le théorème Th.3.3.2 peut être regardé comme un résultat de finitude. En principe, le centre de la grille optimale peut être choisi d'une façon infinie sur l'axe des réels, mais ce théorème assure que l'on ne perd pas en optimisation si le choix se restreint à un ensemble fini, en l'occurrence l'ensemble des points d'entrée de la séquence.

En ce sens, ce théorème se rapproche d'un résultat central de programmation linéaire, i.e. la minimisation d'une fonction linéaire de n variables sur un polyèdre de dimension n : bien que le polyèdre consiste en une infinité de points, il est suffisant de rechercher le point minimum sur l'ensemble fini des sommets du polyèdre [Afriat, 1971], [Afriat, 1973].

Remarque 3 : Nous pouvons interpréter le théorème Th.3.3.2 d'un point de vue de la réalisation musicale : la grille, reliée au tempo intérieur du musicien, reste mathématiquement inaccessible si celui-ci ne joue pas au moins un temps de son tempo intérieur. Ce résultat est important et dépend également de l'hypothèse de perception d'une distance temporelle (ou spatiale) de manière approximée et comparative par notre système psycho-cognitif. Réciproquement, ce théorème permet d'affirmer que si le musicien intériorise une grille de tempo mais ne joue aucun temps sur cette grille (par exemple avec un décalage, anticipation ou retard, couplé ou non à un pattern rythmique), alors la perception du tempo est dépendante d'une information complémentaire culturelle (ou visuelle), mais elle est inaccessible mathématiquement.

Finalement, nous pouvons décrire la procédure pour la résolution de la minimisation unidimensionnelle de (3.3.5)

De la démonstration du Lemme 3.3.1 , on peut écrire

$$G_p(\Delta; q, l) = \sum_{k=1}^n \min_{h=0,1,\dots,p} |t_k - t_q - (h - l)\Delta| \quad (3.3.20)$$

On voit que $G_p(\Delta; q, l)$ est une fonction de Δ linéaire par morceaux avec les points anguleux

$$\gamma_{qlkh} = \frac{t_k - t_q}{h - l} \quad \text{avec } k = 1, \dots, n, h = 0, \dots, p, h \neq l, q \text{ et } l \text{ fixés} \quad (3.3.21)$$

Le minimum de $G_p(\Delta; q, l)$ est atteint en un de ses points anguleux de (3.3.21). La procédure formelle pour résoudre (3.3.5) est alors exposée avec l'algorithme Alg.3.4.

Algorithm 3.4 Meilleur pas

données :

$\{t_1, \dots, t_n\}$, séquence de n points

p , nombre de points de grille en plus du centre

$q \in \{1, \dots, n\}$, indice du point pivot considéré de la séquence

$l \in \{0, \dots, p\}$, nombre de points à gauche de t_q

I , intervalle de faisabilité du pas de grille

1: **for** $k = 1$ to n avec $k \neq q$ **do**

2: **for** $h = 0$ to p avec $h \neq l$ **do**

3:

$$\Delta_{kh} \leftarrow \frac{t_k - t_q}{h - l} \quad (3.3.22)$$

$$\Phi_{kh} \leftarrow \sum_{j=1}^n \min\{\text{frac}(t_j - t_q, \Delta_{kh}), \Delta_{kh} - \text{frac}(t_j - t_q, \Delta_{kh})\} \quad (3.3.23)$$

$$\text{avec } \text{frac}(x, y) = x - \left\lfloor \frac{x}{y} \right\rfloor y$$

4: **end for**

5: **end for**

6: $\bar{h}$ et $\bar{k}$ tels que

$$\Phi_{\bar{h}\bar{k}} \leftarrow \min_{h,k} \{\Phi_{hk}\} \quad (3.3.24)$$

$$\Delta \leftarrow \Delta_{\bar{h}\bar{k}} \quad (3.3.25)$$

7: **return** Δ , taille de pas de la p -grille de meilleure coïncidence centrée sur t_q
et comportant l points à gauche de t_q et $p - l$ points à droite de t_q

Pour chaque q et l donnés, l'algorithme Alg.3.4 est de complexité $\mathcal{O}(n^2p)$ et nécessite np points anguleux différents. Remarquons que l'algorithme de recherche du meilleur pas ne nécessite pas un classement ordonné des points d'entrée mais itère sur les points dans un ordre quelconque, ce qui est très pratique si l'on doit rajouter des points à une séquence donnée. Finalement, la complexité globale de l'algorithme Alg.3.3 est $\mathcal{O}(n^3p^2)$.

3.3.3 Recherche du meilleur nombre de points de grille : indice de coïncidence

Dans les algorithmes précédemment décrits, le nombre $p + 1$ de points de grille a été défini à l'avance, sujet aux bornes dépendantes de la précision désirée, comme en (3.1.26).

Cependant, dans certains cas, le nombre de points de grille peut être diminué sans affecter l'erreur mesurée (3.3.2) comme étant la somme des distances de tous les points d'entrée à leur point de grille le plus proche. Sur la figure Fig.3.14 les points d'entrée sont notés $\bullet$ et les points de grille $\bigcirc$

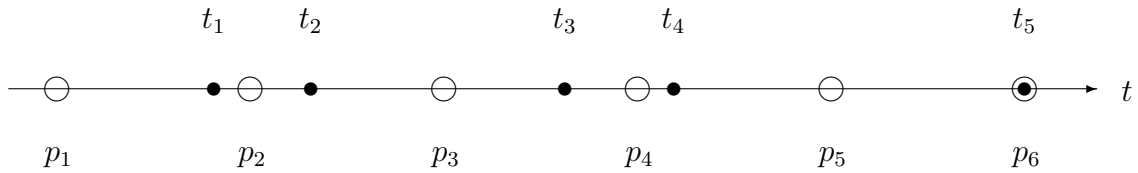


FIG. 3.14 – Recherche du meilleur nombre de points de grille

Dans cet exemple, nous avons une grille optimale de 6 points $\{p_1, \dots, p_6\}$. Si on efface tous les points de grille impairs, la grille résultante $\{p_2, p_4, p_6\}$ est toujours optimale et l'erreur reste la même. $\{p_2, p_4, p_6\}$ est certainement une grille plus efficace que la grille $\{p_1, \dots, p_6\}$, dont les points impairs sont inutiles pour réduire l'erreur.

La discussion ci-dessus montre qu'il est souhaitable d'obtenir une erreur minimale avec un nombre de points de grille le plus petit possible. Pour cela, nous considérons comme

critère de choix du pas Δ de minimiser la fonction ρ d'erreur normalisée suivante :

$$\begin{aligned}\rho(\Delta) &= \frac{f(\Delta)}{\Delta} \quad \text{avec } f \text{ définie en (3.2.10)} \\ &= \frac{1}{\Delta} \left(\sum_{k=1}^n \min_{h \in \mathbb{N}} |t_k - h\Delta| \right) \\ &= \sum_{k=1}^n \min \left\{ \text{frac} \left(\frac{t_k}{\Delta} \right), 1 - \text{frac} \left(\frac{t_k}{\Delta} \right) \right\}\end{aligned}\tag{3.3.26}$$

$$\text{où} \quad \text{frac}(x) = x - \lfloor x \rfloor \quad \text{et} \quad \mathbb{N} = \{0, 1, 2, \dots\}\tag{3.3.27}$$

Pour simplifier, nous venons de considérer et considérons dans le reste de cette section que la grille est centrée sur l'origine (i.e. $s = 0$). Sinon nous aurions à la place de (3.3.26) :

$$\sum_{k=1}^n \min \left\{ \text{frac} \left(\frac{t_k - t_q}{\Delta} \right), 1 - \text{frac} \left(\frac{t_k - t_q}{\Delta} \right) \right\}$$

Notons que, puisque la distance de chaque point t_k à son point de grille le plus proche est au plus $\frac{\Delta}{2}$, l'erreur normalisée prend toujours ses valeurs dans l'intervalle $[0, \frac{n}{2}]$. Dans le but d'obtenir un indicateur, que nous appelons *indice de coïncidence*, à valeur dans $[0, 1]$, il est préférable de considérer la nouvelle fonction d'erreur σ normalisée par $\frac{n}{2}$

$$\sigma(\Delta) = \frac{2f(\Delta)}{n\Delta}\tag{3.3.28}$$

L'indice de coïncidence $\sigma(\Delta)$ présente la propriété de retourner une valeur réelle entre 0 et 1 indépendante de Δ ce qui en fait un paramètre de comparaison de l'ensemble des $\{\Delta\}$. Regardons maintenant de plus près la fonction $\rho(\Delta)$.

Définition 3.3.1. Nous appelons fonction *seawave* une fonction de la forme

$$\begin{aligned}swv_t(x) &= \min_{h \in \mathbb{N}} \left| \frac{t}{x} - h \right| \\ &= \min \left\{ \text{frac} \left(\frac{t}{x} \right), 1 - \text{frac} \left(\frac{t}{x} \right) \right\} \quad , \quad x > 0\end{aligned}\tag{3.3.29}$$

où nous appelons le paramètre $t > 0$ la *phase* de la fonction, et $\mathbb{N} = \{0, 1, 2, \dots\}$

de (3.3.26) , il apparaît que l'erreur normalisée $\rho(\Delta)$ est la somme de n fonctions

seawave avec les phases respectives $t_1, \dots, t_n$

$$\rho(\Delta) = \sum_{k=1}^n \rho_k(\Delta) \quad (3.3.30)$$

avec

$$\rho_k(\Delta) = \text{swv}_{t_k}(\Delta) \quad (3.3.31)$$

La fonction *seawave* $\text{swv}_t(x)$ a la forme suivante :

$$\text{swv}_t(x) = \begin{cases} \frac{t}{x} & , \quad x \geq 2t \\ 1 - \frac{t}{x} & , \quad t \leq x \leq 2t \\ \frac{t}{x} - h & , \quad \frac{t}{h+1} \leq x \leq \frac{2t}{2h+1} \quad (h = 1, 2, \dots) \\ h + 1 - \frac{t}{x} & , \quad \frac{2t}{2h+1} \leq x \leq \frac{t}{h} \quad (h = 1, 2, \dots) \end{cases} \quad (3.3.32)$$

et son graphe est représenté en Fig.3.15.

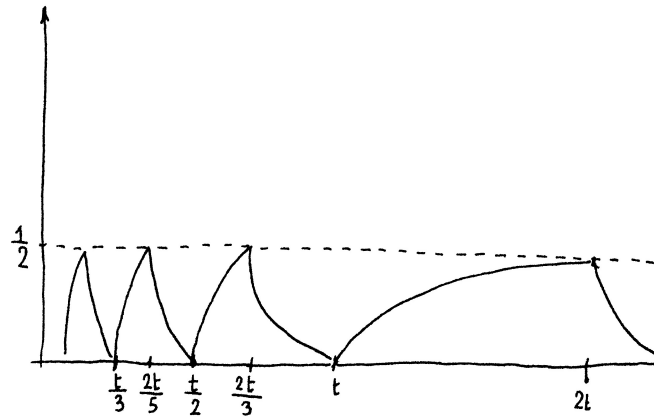


FIG. 3.15 – fonction *seawave*

Remarquons que

- $\text{swv}_t(x)$ est convexe sur les intervalles $[2t, +\infty]$ et $[\frac{t}{h+1}, \frac{2t}{2h+1}]$, $h = \{1, 2, \dots\}$

Sur chacun de ces intervalles, le graphe est un arc d'hyperbole

- $\text{swv}_t(x)$ est concave sur les intervalles $[t, 2t]$ et $[\frac{2t}{2h+1}, \frac{t}{h}]$, $h = \{1, 2, \dots\}$
- $\text{swv}_t(x)$ a un minimum global pour $x = \frac{t}{h}$, $h = \{1, 2, \dots\}$ et un maximum global pour $x = \frac{2t}{2h+1}$, $\{h = 0, 1, \dots\}$. Le minimum global est 0 et le maximum global est $\frac{1}{2}$.

Pour chaque k , tous les points minimum locaux $\frac{t_k}{h}$ de ρ_k sont des points anguleux de

ρ_k et donc de ρ . Dans le cas où $s \neq 0$ les points anguleux sont de la forme $\frac{t_k - t_q}{h - l}$. Est-ce que ρ admet d'autres points minimum locaux? La réponse est affirmative, mais ces minimum locaux sont très particuliers, comme l'affirme le théorème suivant.

Théorème 3.3.6. *Chaque point minimum local de ρ est soit un point minimum global d'un ρ_k (et donc un point anguleux), ou bien il a la propriété que ρ soit constante dans un voisinage ouvert de ce point.*

Démonstration :

Pour chaque $\Delta > 0$, ρ_k prend soit la valeur

$$\frac{t_k}{\Delta} - h, \quad h = \{0, 1, 2, \dots\} \quad (3.3.33)$$

$$\text{ou bien la valeur} \quad h + 1 - \frac{t_k}{\Delta}, \quad h = \{0, 1, 2, \dots\} \quad (3.3.34)$$

Soit $K \equiv K(\Delta)$ l'ensemble des k vérifiant (3.3.33). Supposons que Δ est un minimum local de ρ sans être un point anguleux d'aucun ρ_k . Alors ρ est différentiable en Δ et sa dérivée première est

$$-\frac{1}{\Delta^2} \left(\sum_{k \in K} t_k - \sum_{j \notin K} t_j \right) \quad (3.3.35)$$

et sa dérivée d'ordre i en Δ est

$$(-1)^i \frac{i}{\Delta^{i+1}} \left(\sum_{k \in K} t_k - \sum_{j \notin K} t_j \right) \quad (3.3.36)$$

Puisque Δ doit être un point stationnaire, la dérivée première s'annule en Δ , ce qui implique

$$\sum_{k \in K} t_k - \sum_{j \notin K} t_j = 0 \quad (3.3.37)$$

d'où les dérivées d'ordre supérieure s'annulent également en Δ . ρ est alors constant dans un voisinage ouvert de Δ .

□

Un point Δ tel que ρ est constante dans un voisinage ouvert de Δ a un point anguleux

à la fois à sa droite et à sa gauche :

- à sa droite parce que ρ n'est pas constante pour $\Delta \geq 2 \max_k t_k$;
- à sa gauche parce que pour chaque k on peut toujours trouver h suffisamment grand tel que $\frac{t_k}{h} < \Delta$.

Soit λ et μ respectivement le point anguleux de plus grand h à gauche de Δ , et le point anguleux de plus petit h à droite de Δ . Alors $[\lambda, \mu]$ est le plus grand intervalle fermé contenant Δ où ρ est constante.

On a alors $\rho(\Delta) = \rho(\lambda) = \rho(\mu)$. La figure **Fig.3.16** illustre la fonction $\rho(\Delta)$ somme de quatre fonctions $\rho_k(\Delta)$ pour la séquence [2 3 4 5] qui présente cette particularité.

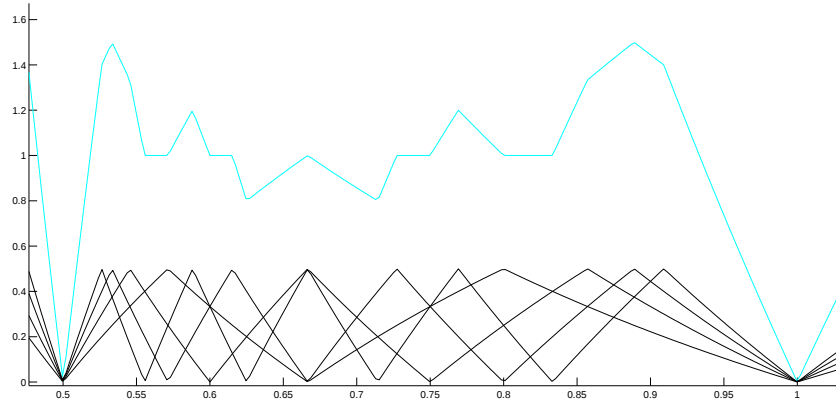


FIG. 3.16 – points anguleux et intervalles constants de ρ avec la séquence [2 3 4 5] et $p = 10$

De la précédente discussion, il ressort que l'on peut toujours restreindre la recherche du point minimum de ρ à l'ensemble des points anguleux ($\frac{t_k}{h}$ si $s = 0$ ou $(\frac{t_k - t_q}{h - l})$ dans le cas général) de ρ .

Remarquons que ces points anguleux coïncident avec ceux de f .

Nous pouvons conclure avec la description de l'algorithme **Alg.3.3.3** permettant de minimiser ρ .

La complexité de cet algorithme est $O(n^2 p_{max})$.

Algorithm 3.5 meilleur pas avec erreur normalisée**données**

$\{t_1, \dots, t_n\}$, séquence de n points
 c , borne inférieure pour Δ
 $q \in \{1, \dots, n\}$, indice du point pivot considéré de la séquence
 $l \in \{0, \dots, p\}$, nombre de points à gauche de t_q
 I , intervalle de faisabilité du pas de grille

$p_{max} \leftarrow 1 + \lfloor \frac{t_n}{c} \rfloor$, d'après (3.1.26)

for $k = 1$ to n avec $k \neq q$ **do**

for $h = 0$ to p_{max} avec $h \neq l$ **do**

$$\Delta_{kh} \leftarrow \frac{t_k - t_q}{h - l}$$

$$\rho_{kh} \leftarrow \frac{f(\Delta_{kh})}{\Delta_{kh}}$$

end for

end for

k^* et h^* tels que $\rho_{k^*h^*} \leftarrow \min_{k,h} \rho_{kh}$

$$\Delta \leftarrow \Delta_{k^*h^*}$$

$$p \leftarrow \left\lceil \frac{t_{k^*}}{\Delta} \right\rceil$$

return Δ, p

3.3.4 Réduction de l'algorithme d'optimisation

L'algorithme Alg.3.4 retourne, pour q et l donnés, la meilleure p -grille centrée sur t_q . Pour cela les candidats $\frac{t_k - t_q}{h - l}$ sont itérés sur k, h et l avec $1 \leq k \leq n$, $0 \leq l \leq p$ et $0 \leq h \leq p$, $h \neq l$. Nous pouvons alors éliminer les doublons en parcourant uniquement l'ensemble $\frac{t_k - t_q}{u}$ avec $-p \leq u \leq p$ que nous pouvons réduire à $0 \leq u \leq p$ en considérant la symétrie induite par le changement de signe qui concerne l'ordre inverse des points. La complexité de l'algorithme est ainsi abaissée à $\mathcal{O}(\frac{n^2(n-1)p}{2})$ au lieu de $\mathcal{O}(n^3p^2)$ si l'on considère que l'on perd maintenant l'information l qui n'est pas très nécessaire pour la solution de notre problème. Nous pouvons proposer l'algorithme réduit final Alg.3.6.

Quelques remarques sur l'algorithme réduit

La simplicité de la forme de Alg.3.6 montre comment considérer les intervalles inter-onsets : c'est parmi leurs sous-multiples que se trouve la valeur-intervalle qui permet

Algorithm 3.6 Meilleure grille d'approximation bidimensionnelle**données** $\{t_1, \dots, t_n\}$, séquence de n points p , nombre de points de grille en plus du centre**for** $q = 1$ to n **do** **for** $k = q + 1$ to n **do** **for** $h = 0$ to p **do**

$$\Delta_{qkh} \leftarrow \frac{t_k - t_q}{h}$$

$$\Phi_{qkh} \leftarrow \sum_{j=1}^n \min \left\{ \text{frac} \left(\frac{t_j - t_q}{\Delta_{qkh}} \right), 1 - \text{frac} \left(\frac{t_j - t_q}{\Delta_{qkh}} \right) \right\} \text{ avec } \text{frac}(x) = x - \lfloor x \rfloor$$

end for **end for****end for** q^*, k^* et h^* tels que $\Phi_{q^*k^*h^*} \leftarrow \min_{q,k,h} \Phi_{qkh}$ $\Delta^* \leftarrow \Delta_{q^*k^*h^*}$ **return** q^*, Δ^* , centre t_{q^*} et pas de la grille de coïncidence optimale Δ^*

la minimisation de la fonction d'erreur de quantification. Le paramètre p apparaît ici clairement comme un élément de *résolution temporelle de quantification*, alors même qu'il est défini comme la taille de la grille à *comparer* avec la séquence d'entrée. En d'autres termes, nous considérons sans répétition tous les inter-onsets positifs entre chaque paire de points de la séquence, puis leur sous multiples d'ordre 1 à p : l'exploration exhaustive de cet ensemble discret aboutit au minimum, sur la surface continue de définition de la fonction d'erreur, de la recherche optimale de la solution Δ^* .

Il est intéressant de noter également que l'on explore avec **Alg.3.6** seulement les $n - 1$ premiers points en tant que *points pivots*, i.e. points susceptibles d'être le centre de la grille. Si l'on considère deux points, il est évident que l'on peut prendre comme centre chacun des deux points du fait de leur symétrie . On peut démontrer intuitivement que pour un ensemble de n points, le centre de la p -grille optimale se trouve nécessairement parmi $n - 1$ points constitués par les $n - 2$ points compris entre les deux extrémités, et l'extrémité gauche ou l'extrémité droite. Nous généraliserons cette idée ci-après sous la forme d'un second théorème.

3.3.5 La grille optimale contient deux points d'entrée

Nous établissons maintenant un second théorème sur la grille optimale et démontrons que le théorème Th.3.2.4 peut être encore étendu. Nous montrons que la grille de coïncidence optimale a non seulement pour centre un point d'entrée, mais contient également un second point d'entrée, ce qui implique finalement que toute grille optimale a deux centres distincts.

Nous avons montré que, si (s^*, Δ^*) minimise la fonction d'erreur

$$f(s, \Delta) = \sum_{k=1}^n \min_{h=0, \dots, p} |t_k - (s + h\Delta)| \quad (3.3.38)$$

alors Δ^* se trouve parmi les Δ qui vérifient $s^* + h\Delta = t_k$ pour un couple (h, k) . Ainsi, sans perte de généralité, nous pouvons supposer que $s^* = 0 = t_m$, d'où la configuration suivante :

$$t_1 < \dots < t_{m-1} < 0 \equiv t_m < t_{m+1} < \dots < t_n$$

et les p points de grille sont

$$-\nu\Delta, -(\nu-1)\Delta, \dots, -\Delta, 0, \Delta, 2\Delta, \dots, \mu\Delta$$

avec $p = \mu + \nu$. La fonction d'erreur $f(s, \Delta)$ prend alors la forme

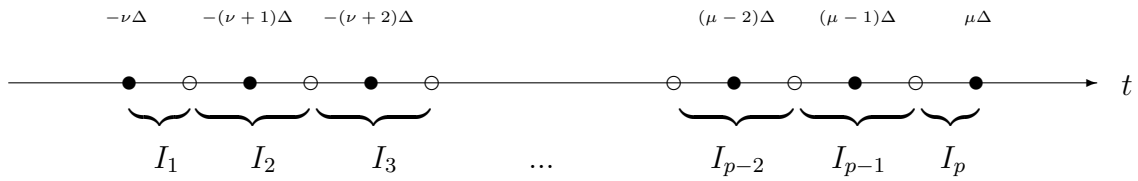
$$f(\Delta) \equiv f(0, \Delta) = \sum_{k=1, k \neq m}^n \min_{h=-\nu, \dots, \mu} |t_k - h\Delta| \quad (3.3.39)$$

Au regard de (3.1.15) et de (3.1.21), la fonction est linéaire continue par morceaux et sa pente n'est jamais nulle, d'où nous pouvons établir la propriété suivante :

Lemme 3.3.7. *Il n'y a pas d'intervalle ouvert de longueur positive où $f(\Delta)$ est constante.*

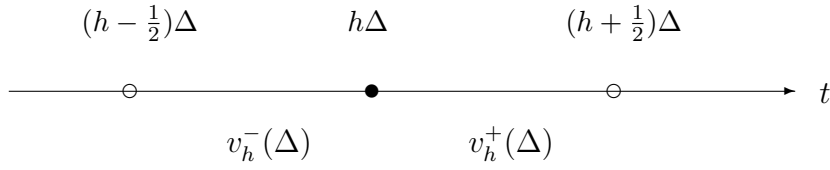
Maintenant, partitionons l'intervalle $[-\nu\Delta, \mu\Delta]$ en p sous-intervalles (tous semi-ouverts sauf I_1) :

$$\begin{aligned}
I_1 &= \left[-\nu\Delta, -(\nu - \frac{1}{2})\Delta \right] \\
I_2 &= \left[-(\nu - \frac{1}{2})\Delta, -(\nu - \frac{3}{2})\Delta \right] \\
I_3 &= \left[-(\nu - \frac{3}{2})\Delta, -(\nu - \frac{5}{2})\Delta \right] \\
&\dots \\
I_{p-2} &= \left[(\mu - \frac{5}{2})\Delta, (\mu - \frac{3}{2})\Delta \right] \\
I_{p-1} &= \left[(\mu - \frac{3}{2})\Delta, (\mu - \frac{1}{2})\Delta \right] \\
I_p &= \left[(\mu - \frac{1}{2})\Delta, \mu\Delta \right]
\end{aligned}$$



- Pour tout point $t_k \in I_1$, le point de grille le plus proche est $-\nu\Delta$
- Pour tout point $t_k \in I_2$, le point de grille le plus proche est $-(\nu - 1)\Delta$
- ...
- Pour tout point $t_k \in](h - \frac{1}{2})\Delta, (h + \frac{1}{2})\Delta]$, le point de grille le plus proche est $h\Delta$
- ...
- Pour tout point $t_k \in I_{p-1}$, le point de grille le plus proche est $(\mu - 1)\Delta$
- Finalement, pour tout point $t_k \in I_p$, le point de grille le plus proche est $\mu\Delta$

Partitionnons chaque intervalle $] (h - \frac{1}{2})\Delta, (h + \frac{1}{2})\Delta]$ autre que I_1 et I_p en deux demi-intervalles $] (h - \frac{1}{2})\Delta, h\Delta]$ et $] h\Delta, (h + \frac{1}{2})\Delta]$, et posons $v_h^-(\Delta)$ et $v_h^+(\Delta)$ compteurs du nombre de points t_k respectivement dans le premier demi-intervalle puis le second.



Remarquons que le demi-intervalle gauche est vide pour I_1 et le demi-intervalle droit est vide pour I_p . Remarquons également que chaque point de l'intervalle $] -\nu\Delta, \mu\Delta]$ a un point de grille le plus proche, excepté les *points-milieu* $m_h = (h + \frac{1}{2})\Delta$ qui sont équidistants de deux points de grille : $h\Delta$ et $(h + 1)\Delta$

Lemme 3.3.8. *Pour un ε suffisamment petit, si chaque point $t_k \in](h - \frac{1}{2})\Delta, (h + \frac{1}{2})\Delta]$ n'est ni un point de grille ou un point-milieu – i.e. différent de $(h - \frac{1}{2})\Delta$, $h\Delta$, $(h + \frac{1}{2})\Delta$ – alors*

$$v_h^-(\Delta - \varepsilon) = v_h^-(\Delta + \varepsilon) = v_h^-(\Delta) \quad (3.3.40)$$

$$v_h^+(\Delta - \varepsilon) = v_h^+(\Delta + \varepsilon) = v_h^+(\Delta) \quad (3.3.41)$$

Remarquons que Lem.3.3.8 décrit l'existence d'un état invariant par une homothétie infinitésimale de la grille.

Théorème 3.3.9. *Si Δ^* minimise $f(\Delta)$, alors un point $t_k \neq 0$ doit être un point de la grille $h\Delta^*$*

Démonstration :

Nous distinguons 3 cas :

1. aucun point t_k ne coïncide avec un point de grille ou un *point-milieu*
2. il existe un point t_k qui coïncide avec un *point-milieu*, mais aucun autre point t'_k ne coïncide avec un point de la grille
3. il existe un point t_k qui coïncide avec un point de la grille

Nous allons montrer que les deux premiers cas sont en contradiction avec le caractère optimal de Δ^* . Puisque les 3 cas sont exhaustifs et mutuellement exclusifs, le seul cas possible est le troisième cas, ce qui établit notre thèse.

1. Supposons que nous sommes dans le premier cas. Nous avons alors de par le

Lem.3.3.8 :

$$f(\Delta^* + \varepsilon) - f(\Delta^*) = \sum_{h=-\nu}^{h=\mu} h (v_h^-(\Delta^*) - v_h^+(\Delta^*)) \varepsilon \equiv V\varepsilon \quad (3.3.42)$$

$$f(\Delta^* - \varepsilon) - f(\Delta^*) = \sum_{h=-\nu}^{h=\mu} h (v_h^+(\Delta^*) - v_h^-(\Delta^*)) \varepsilon \equiv -V\varepsilon \quad (3.3.43)$$

En effet, expliquons (3.3.42) :

si $t_k \in \left] h_k \Delta^*, (h_k + \frac{1}{2}) \Delta^* \right[$
 $|t_k - h_k \Delta^*|$ devient $|t_k - h_k(\Delta^* + \varepsilon)| = |t_k - h_k \Delta^*| - h_k \varepsilon$
 Comme il y a $v_{h_k}^+(\Delta^*)$ points dans l'intervalle $\left] h_k \Delta^*, (h_k + \frac{1}{2}) \Delta^* \right[$, leurs contributions à $f(\Delta^* + \varepsilon) - f(\Delta^*)$ est de $-h_k \varepsilon v_{h_k}^+(\Delta^*)$

Si $t_k \in \left] (h_k - \frac{1}{2}) \Delta^*, h_k \Delta^* \right[$
 $|t_k - h_k \Delta^*|$ devient $|t_k - h_k(\Delta^* + \varepsilon)| = |t_k - h_k \Delta^*| + h_k \varepsilon$.
 Il y a $v_{h_k}^-(\Delta^*)$ points dans l'intervalle $\left] (h_k - \frac{1}{2}) \Delta^*, h_k \Delta^* \right[$, leurs contributions à $f(\Delta^* + \varepsilon) - f(\Delta^*)$ est de $h_k \varepsilon v_{h_k}^-(\Delta^*)$

Nous obtenons (3.3.43) par le même raisonnement.

Au regard de (3.3.42) et (3.3.43), V ne peut pas être nul, sinon Lem.3.3.7 serait contredit. Nous devons alors avoir soit $f(\Delta^* + \varepsilon) < f(\Delta^*)$ ou $f(\Delta^* - \varepsilon) < f(\Delta^*)$. Dans les deux cas, il y a contradiction avec l'optimalité de Δ^* .

2. Supposons que nous sommes dans le second cas. Soit, pour chaque point t_k , $h_k \Delta$ le point de grille le plus proche (si $t_k = (h + \frac{1}{2}) \Delta$ est un *point-milieu*, nous attribuons t_k à $h \Delta$). Soit $\mathcal{M}$ l'ensemble des points d'entrée qui sont des *points-milieu* et $\mathcal{N}$ l'ensemble des points restants. On a alors

$$f(\Delta^*) = \sum_{k \in \mathcal{M}} |t_k - h_k \Delta^*| + \sum_{k \in \mathcal{N}} |t_k - h_k \Delta^*| \quad (3.3.44)$$

Examinons l'effet sur $f(\Delta)$ de l'accroissement de Δ de Δ^* à $\Delta^* + \varepsilon$.

Pour $k \in \mathcal{M}$, puisque t_k est équidistant de $h_k \Delta^*$ et de $(h_k + 1) \Delta^*$, sa distance de $h_k(\Delta^* + \varepsilon)$ diminuerait de $h_k \varepsilon$, et celle de $(h_k + 1)(\Delta^* + \varepsilon)$ augmenterait de $(h_k + 1) \varepsilon$. Donc sa contribution à $f(\Delta^* + \varepsilon) - f(\Delta^*)$ serait négative, et le point t_k passe dans l'ensemble comptabilisé par $v_h^-(\Delta^* + \varepsilon) = v_h^-(\Delta^*)$.

De l'autre coté, si Δ^* diminue jusque $\Delta^* - \varepsilon$, la distance de t_k , $k \in \mathcal{M}$, de $h_k(\Delta^* - \varepsilon)$ augmenterait de $h_k\varepsilon$ et sa distance de $(h_k + 1)(\Delta^* - \varepsilon)$ diminuerait de $(h_k + 1)\Delta^*$, et alors la contribution de t_k à $f(\Delta^* + \varepsilon) - f(\Delta^*)$ serait négative, et le point t_k passe dans l'ensemble comptabilisé par $v_h^+(\Delta^* - \varepsilon) = v_h^+(\Delta^*)$.

Les points d'entrée t_k restants, $k \in \mathcal{N}$ reviennent au premier cas. Donc si nous écrivons :

$$f_{\mathcal{M}}(\Delta) = \sum_{k \in \mathcal{M}} \min_{h=-\nu, \dots, \mu} |t_k - h\Delta| \quad (3.3.45)$$

$$f_{\mathcal{N}}(\Delta) = \sum_{k \in \mathcal{N}} \min_{h=-\nu, \dots, \mu} |t_k - h\Delta| \quad (3.3.46)$$

nous obtenons :

$$f(\Delta^* + \varepsilon) - f(\Delta^*) = (f_{\mathcal{M}}(\Delta^* + \varepsilon) - f_{\mathcal{M}}(\Delta^*)) + (f_{\mathcal{N}}(\Delta^* + \varepsilon) - f_{\mathcal{N}}(\Delta^*)) \quad (3.3.47)$$

$$f(\Delta^* - \varepsilon) - f(\Delta^*) = (f_{\mathcal{M}}(\Delta^* - \varepsilon) - f_{\mathcal{M}}(\Delta^*)) + (f_{\mathcal{N}}(\Delta^* - \varepsilon) - f_{\mathcal{N}}(\Delta^*)) \quad (3.3.48)$$

Nous avons montré que la première parenthèse – de la partie droite – est négative pour (3.3.47) et (3.3.48) et que la seconde parenthèse est de signe opposé dans (3.3.47) et (3.3.48). D'où également dans ce second cas, soit $f(\Delta^* + \varepsilon) < f(\Delta^*)$, soit $f(\Delta^* - \varepsilon) < f(\Delta^*)$, ce qui est en contradiction avec l'optimalité de Δ^* . $\square$

Le théorème **Th.3.3.9** décrit simplement que la grille de coïncidence optimale passe par deux points donnés. En fait, le mode de construction des candidats Δ_{qlkh} (3.3.22) de l'algorithme **Alg.3.4** utilise déjà implicitement cette propriété avec (3.3.21) sans qu'elle ait été démontrée auparavant. C'est (3.3.20) qui suggère cette résolution des points anguleux ou singuliers en référence à la résolution géométrique établie en (3.1.21) pour trouver les solutions α_{kh} du problème initial et simplifié (3.1.3).

De même, le théorème **Th.3.3.9** étend la remarque 3 (3.3.2) en assurant que si une séquence homorythmique approche de façon optimale une phrase d'événements temporels, alors deux de ces événements tombent sur les temps de la séquence de pulsation.

Il est intéressant de voir que le fait de trouver ainsi deux points de synchronisation permet d'établir et d'affirmer une propriété de périodicité. C'est la même notion

de périodicité qui est implicitement suggérée par d’autres travaux qui utilisent des techniques différentes. Ainsi la transformée en ondelettes utilisée dans les travaux de [Smith, 2000], [Smith and Honing, 2008] va chercher une continuité de la périodicité vers une fréquence “fondamentale” du signal des impulsions rythmiques : le fait de trouver un résultat et de le lire graphiquement sur le spectre en ondelettes met en évidence la périodicité : la répétition qui a besoin de deux points pour exister. La périodicité qui peut construire une phrase rythmique va chercher ses origines vers une périodicité élémentaire des constituants de la phrase rythmique, et quelle périodicité plus simple y-a-t-il que de regarder les intervalles constitués par chaque couple de points et leurs sous-multiples ? Seppanen [2001] a également cette même intuition, mais ne considère qu’un sous-ensemble des solutions de périodicité dans la recherche du *tactus*. Le théorème Th.3.3.9 est une généralisation de la périodicité fractale qui laisse émerger à partir de petits intervalles de temps la structure de la séquence donnée. Ainsi, même si nous ne considérons pas de hiérarchie de tempo ou de métrique, ce mécanisme *bottom-up* pourrait peut-être expliquer la perception de différents *tempi* subjectifs d’une même séquence d’événements temporels.

Chapitre 4

Accélération de tempo : extension de la grille optimale

Nous abordons dans ce chapitre la problématique d'un tempo accéléré uniformément et continûment. Pour cela, nous étendons les recherches de [Guerino Mazzola \[1994\]](#) sur la notion de courbe de tempo. Nous introduisons une formulation simple de deux opérateurs caractérisant une transformation temporelle par accélération directe uniforme (4.1.3) et inversement une transformation temporelle par accélération uniforme réciproque (4.1.4). Nous intégrons ces résultats dans le modèle de *grille optimale bidimensionnelle* où un nouveau paramètre d'accélération θ entraîne une minimisation d'une fonction dans un espace tridimensionnel. Nous développons une nouvelle méthode pour approcher la meilleure solution (s, Δ, θ) et donnons, en (4.3), l'algorithme final de la *grille optimale* d'un tempo modélisé par une accélération uniforme. Nous expliquons alors l'utilisation de ces résultats et comment ils conduisent à l'inférence du tempo, détaillée au chapitre (5).

4.1 Tempo uniformément et continûment accéléré

La recherche de la grille de tactus optimale présentée précédemment suppose un tempo constant : la grille optimale comprend un *pas* qui ne varie pas avec le temps. Un tel tempo stable n'est cependant jamais possible dans la réalité, sauf s'il s'agit d'un séquenceur logiciel ! Néanmoins, nous pouvons dans de nombreuses situations considérer que le tempo est *continu*, mais même si cette continuité existe elle ne peut être qu'une modélisation mathématique. La manifestation du tempo est nécessairement discrète, par essence même du rythme généré et par la nature discrète des événements qui le composent. Le tempo apparaît comme un striage ou un découpage particulier du temps musical. Nous considérons que les événements rythmiques *échantillonnent* la courbe continue du tempo et constituent l'unique information à laquelle nous avons accès dans le cas le plus élémentaire. Nous verrons aussi pourquoi une séquence d'événements donne lieu à plusieurs interprétations concurrentes pour l'organisation des valeurs rythmiques et du ou des tempi qui la traversent. Dans ce chapitre, nous proposons une modélisation continue du tempo, et regardons quelle est son implication sur l'algorithme de résolution optimale bidimensionnelle de grille établie précédemment.

4.1.1 Notation et interprétation, transcription et performance : la question du temps en musique

Nous distinguons ici la *partition* de sa *performance* ou *interprétation*. Nous pouvons appréhender le temps musical comme la réalisation de deux temps parallèles : un temps abstrait qui permet la notation sous la forme d'une partition, et le *temps réel* d'une exécution qui permet de saisir et d'écouter une interprétation. Nous montrons ci-après que le temps réel et le temps abstrait sous forme de transcription ou de notation métrée s'articulent réciproquement autour d'une fonction qui définit la notion de *courbe de tempo*. De ce point de vue mathématique, le tempo est donc lié au moins implicitement à une forme de transcription métrée, c'est-à-dire de représentation abstraite d'une mesure du temps musical.

Comme le décrivent [Guerino Mazzola \[1994\]](#), nous nous attachons au passage du temps symbolique, temps *écrit* de la *partition* exprimé en unités de pulsation, au temps réel de sa *réalisation* associé aux secondes qui s'égrainent. Dans le cadre du *suivi de tempo* appliqué aux musiques improvisées, nous ne connaissons pas la partition, parce qu'elle

n'existe peut-être pas avant d'être interprétée comme c'est le cas pour une improvisation totale. Nous supposons que la partition représente l'information symbolique à laquelle nous n'avons pas accès, seules les observations de sa réalisation sont disponibles au fur et à mesure. Nous faisons donc l'hypothèse qu'une forme de représentation musicale existe et qu'elle constitue la partition des événements musicaux. L'interprétation rythmique par le musicien consiste à passer du temps symbolique de la partition au temps physique de sa réalisation. Nous notons d le temps symbolique qui a pour unité la pulsation, et t le temps réel qui a pour unité la seconde. D'après [Guerino Mazzola, 1994], nous modélisons l'interprétation rythmique comme le passage de l'un à l'autre par une fonction cachée f que nous appelons *fonction de performance* ou *fonction d'interprétation* :

$$t = f(d) \quad (4.1.1)$$

Nous définissons alors dans le domaine *pulsationnel* ou symbolique la fonction t' dérivée de t :

$$t' = \frac{\partial t}{\partial d} = \frac{\partial f}{\partial d} \quad (4.1.2)$$

Si f^{-1} désigne la fonction réciproque de f croissante monotone – car la flèche du temps n'est pas inversible – f^{-1} existe et

$$d = f^{-1}(t) \quad (4.1.3)$$

On peut interpréter f^{-1} comme une fonction de *transcription* d'une réalisation temporelle vers la notation symbolique, nous posons alors dans le domaine temporel

$$d' = \frac{\partial f^{-1}}{\partial t} = \frac{\partial d}{\partial t} = \frac{1}{\frac{\partial t}{\partial d}} = \frac{1}{\frac{\partial f}{\partial d}} \quad (4.1.4)$$

Remarque Nous exprimons habituellement le tempo en pulsations par minute – noté *bpm* pour *beat per minutes* – ou en pulsations par seconde que nous notons *bps*. Du point de vue de la notation musicale, on considère la notion de courbe de tempo comme une fonction du temps symbolique d dans lui-même. Ainsi, on notera $T(d_d)$ le tempo dans ce domaine : à telle mesure de la partition le tempo change de valeur, par exemple 72 à la noire mais la durée de la noire reste toujours la même valeur symbolique, on parle aussi d'équivalence métrique. Du point de vue interprétatif ou exécutif, le tempo se manifeste dans le système de coordonnées temps-pulsation (t, d)

– qui rend compte de l’acte de transcription. En effet, c’est bien dynamiquement que la notion de tempo apparaît, comme la variation de la pulsation – unité de la partition – par rapport au temps réel de la réalisation. Par passage à la limite, le tempo est la dérivée de la pulsation par rapport au temps, le tempo est une vitesse :

$$T(t_d) = \frac{\partial d}{\partial t} \quad (4.1.5)$$

$T(t_d)$ désigne le tempo à l’instant t de la réalisation associé à l’instant d de la partition, $T(t_d)$ est une fonction du temps réel. Nous montrons ci-après que l’on peut définir également $T(d_t)$ dans le domaine réciproque – système de coordonnées pulsation-temps (d, t) – le tempo à l’instant d de la partition associé à son temps de réalisation t , le tempo $T(d_t)$ étant une fonction du temps symbolique d de la partition.

On obtient alors avec (4.1.4) et (4.1.5)

$$T(d_t) = \left(\frac{\partial f}{\partial d} \right)^{-1} \quad (4.1.6)$$

ainsi

$$T(d_d) \equiv T(t_d) \equiv T(d_t) \quad \text{mais} \quad T(d_d) \neq T(t_d) \neq T(d_t) \quad (4.1.7)$$

Dans (4.1.7), $T(d_d)$ désigne le tempo dans le temps abstrait de la partition, hors de toute référence à sa réalisation. $T(d_d)$ reste une pure notation – terme que nous préférons ici à celui de transcription qui indiquerait qu’il vient d’une réalisation. $T(t_t)$ indiquerait un tempo dans le temps réel pur sans référence à une quelconque notion de transcription, ce qui revient à nier l’existence d’un tempo et n’a donc pas de sens par rapport à une définition du tempo qui présuppose une forme de transcription ou de notation. $T(t_t)$ est par là même absurde, car t désigne le temps – la flèche du temps irréversible – qu’on ne peut pas abstraire, et qui est paradoxalement dual ou réciproque du temps d qui l’abstrait. Ainsi, la représentation passe par une forme d’abstraction de la réalité.

Pour illustrer les domaines de temps parallèles, quand nous lisons un tempo 72 à la noire dans la partition nous parlons explicitement de $T(d_d)$, nous pensons ou interprétons $T(d_t)$ (4.1.6) – le tempo à l’instant d de la partition que nous sommes en train d’interpréter au temps t , alors que nous générons ou transcrivons $T(t_d)$ (4.1.5) – le tempo à l’instant t de l’interprétation dans le temps transcrit d de la partition. Cette *confusion* ou *coïncidence* entre d_d , d_t et t_d , ou mieux cette équivalence de la pulsation

d au temps d de la partition, du temps t de la pulsation d , et de la pulsation d du temps t vient certainement de notre capacité à abstraire le temps physique musical en une division de pulsations symboliques dont nous percevons mieux la lisibilité que si nous devions la mesurer réellement. Cela rejoint le fait que nos critères d'évaluation ou de mesure soient, pour le domaine temporel, relatifs ou comparatifs, et plus difficilement absolus. La relation (4.1.7) entre $T(t_d)$ et $T(d_t)$ semble tautologique, elle représente la dualité *temps symbolique / temps physique* fondamentale qui met en évidence la non-unicité de la *transcription rythmique* : celle-ci est fonction de l'articulation du tempo dans les domaines symbolique et réel, couplée à un choix de valeurs rythmiques, cette idée est également partagée par [Cemgil, 2004].

Nous tirons alors de (4.1.6) reformulé avec $\frac{\partial f}{\partial d} = \frac{1}{T(d)}$:

$$t = f(d) = \int_{d_0}^d \frac{1}{T(d)} \partial d \quad (4.1.8)$$

avec d_0 la pulsation initiale. $T(d)$ est la *courbe de tempo* dans la partition, une fonction du temps symbolique d en pulsations, par exemple une indication *Accelerando*. Connaissant $T(d)$, nous déduisons la fonction de performance f associée.

(4.1.8) exprime par l'intermédiaire de la fonction de tempo T le glissement du temps symbolique par rapport au temps réel, et par là une relation de phase entre ces deux temporalités duales, une phase des temps. La fonction de tempo T rend explicite cette relation de phase.

4.1.2 Modélisation du tempo par une pulsation accélérée uniformément

(4.1.8) décrit par une intégrale la relation entre la flèche du temps réel et son espace dual monodimensionnel du temps symbolique. Afin de prendre en compte un tempo non constant dans notre *suivi de tempo*, nous modélisons le tempo par cette relation. Si nous pouvons passer facilement d'un espace de temps à un autre, nous trouvons alors une solution pour la transcription.

Nous désirons trouver une expression qui traduise la transformation d'une séquence temporelle d'événements $\{t_n\}$ caractérisée de la manière suivante : $\{t_n\}$ part d'un tempo initial et sa pulsation est linéairement accélérée sur une suite d'intervalles finis

et consécutifs du domaine symbolique, nous avons alors :

$$\frac{\partial T}{\partial d} = \varepsilon \quad (4.1.9)$$

Avec ε constante en “beat” par seconde et par pulsation (*bpspp*). $T(d)$ est une fonction continue par morceaux ou polygonale. Si T_0 est le tempo initial au temps $d_0 = 0$, (4.1.9) signifie qu’on incrémente le tempo de la valeur ε à chaque pulsation :

$$T(d) = T_0 + \varepsilon d \quad (4.1.10)$$

On intègre alors (4.1.8) avec (4.1.22) pour trouver la nouvelle séquence accélérée que nous notons $\{t_n^{(\varepsilon)}\}$ ¹

$$t^{(\varepsilon)} = f(d) = \frac{1}{\varepsilon} \ln\left(1 + \frac{\varepsilon}{T_0} d\right) \quad (4.1.11)$$

$$t^{(\varepsilon)} = \frac{d}{T_0} \left[\frac{\ln\left(1 + \frac{d}{T_0} \varepsilon\right)}{\frac{d}{T_0} \varepsilon} \right] \quad (4.1.12)$$

$$t^{(\varepsilon)} = t \left[\frac{\ln(1 + t\varepsilon)}{t\varepsilon} \right] \quad (4.1.13)$$

En effet, si on considère une séquence de pulsations d_n à tempo constant T_0 , son interprétation temporelle mécanique isochrone est bien t_n avec $t_n = \frac{d_n}{T_0}$ et $t = \frac{d}{T_0}$ représente le temps correspondant à la pulsation à tempo constant T_0 . Nous vérifions également que :

$$\lim_{\varepsilon \rightarrow 0} t^{(\varepsilon)} = t \quad (4.1.14)$$

Ainsi (4.1.13) et (4.1.14) permettent de trouver, à partir d’une séquence temporelle d’événements à tempo constant, la nouvelle séquence temporelle accélérée de ε pulsations par seconde et par pulsation, i.e. le tempo augmente de la même valeur à chaque pulsation.

¹(ε) ne doit pas être compris comme un exposant mais comme la référence à l’incrément linéaire du tempo de la séquence considérée.

En inversant (4.1.11), on obtient

$$d(t) = T_0 t^{(\varepsilon)} \left[\frac{e^{\varepsilon t^{(\varepsilon)}} - 1}{\varepsilon t^{(\varepsilon)}} \right] \quad (4.1.15)$$

$$\text{soit } t = t^{(\varepsilon)} \left[\frac{e^{\varepsilon t^{(\varepsilon)}} - 1}{\varepsilon t^{(\varepsilon)}} \right] \quad (4.1.16)$$

$$\text{et par passage à la limite } \lim_{\varepsilon \rightarrow 0} t = t^{(\varepsilon)} \quad (4.1.17)$$

Quelle est la signification de (4.1.16) et (4.1.17) ? Étant donnée une séquence temporelle $\{t_n^{(\varepsilon)}\}$ que l'on suppose uniformément accélérée de ε pps par pulsation à partir d'un tempo initial T_0 , on retrouve la séquence temporelle non accélérée $\{t_n\}$ qui correspond au tempo constant initial T_0 .

Il est intéressant de noter que l'équation (4.1.13) et que son expression inverse (4.1.16) sont indépendantes de T_0 : cela signifie que l'on peut accélérer de façon directe (4.1.13) ou inverse (4.1.16) une séquence temporelle sans connaître au préalable son tempo de référence interne qui décrit son organisation rythmique à l'aide de ses seules proportions. Une justification mathématique est que l'on reste dans le domaine temporel sans avoir besoin du domaine symbolique pour effectuer cette transformation.

Finalement (4.1.22) et (4.1.16) donnent :

$$T(t^{(\varepsilon)}) = T_0 e^{\varepsilon t^{(\varepsilon)}} \quad (4.1.18)$$

(4.1.18) montre que si la courbe de tempo est une droite dans la partition, elle est exponentielle dans le domaine temporel associé à sa réalisation. Les fonctions (4.1.13) et (4.1.16) permettent de passer d'un domaine à un autre.

On pourrait modéliser la courbe de tempo dans la partition par une autre fonction, logarithmique, exponentielle ou encore différente. Nous manquons ici d'éléments psychoacoustiques sur l'évolution de la courbe de tempo dans un contexte musical particulier. Drake and Penel [1998] mettent en évidence les sources principales de la variation de tempo à l'aide d'une hypothèse perceptuelle sur un modèle psychologique d'organisation temporelle des événements : extraction de la régularité et segmentation en

phrases musicales. Cependant, les auteurs n'indiquent pas une description générale de la forme de la variation du tempo quand celle-ci est intentionnelle ou pas. Il nous faudrait mener ici une autre étude psychoacoustique sur le mécanisme d'interpolation de tempi, sur une base de musiciens de styles musicaux différents. Pour le moment, nous choisissons quand même une modélisation linéaire du tempo dans le domaine du temps symbolique de la partition. L'idée qui nous guide dans ce choix est que localement, si le tempo ne varie pas de manière trop brusque, on peut toujours faire l'approximation de son évolution par un ensemble de segments de droite.

Nous récapitulons ci-dessous, dans le domaine du temps de la performance, les formules d'accélération directe uniforme, d'accélération réciproque uniforme et de courbe de tempo accélérée uniformément en fonction de l'incrément de tempo ε :

$$\text{Accélération uniforme directe : } t^{(\varepsilon)} = \frac{\ln(1 + t\varepsilon)}{\varepsilon} \quad (4.1.19)$$

$$\text{Accélération uniforme réciproque : } t = \frac{e^{\varepsilon t^{(\varepsilon)}} - 1}{\varepsilon} \quad (4.1.20)$$

$$\text{Courbe de tempo accéléré uniformément : } T(t^{(\varepsilon)}) = T_0 e^{\varepsilon t^{(\varepsilon)}} \quad (4.1.21)$$

$$T(t) = T_0 + \varepsilon t \quad (4.1.22)$$

4.1.3 Accélération directe uniforme

À partir des résultats précédemment obtenus pour la modélisation du tempo accéléré uniformément, nous désirons trouver sous une autre forme l'image accélérée d'une séquence temporelle donnée, et par la suite l'image accélérée réciproquement d'une séquence accélérée, c'est-à-dire la séquence non accélérée.

Soit $\{t_0, t_1, \dots, t_n\}$ une séquence de $n+1$ points. On désire trouver la séquence $\{t_i^{(\varepsilon)}\}$ correspondante accélérée d'un facteur α donné, soit $\{t_0^{(\varepsilon)}, t_1^{(\varepsilon)}, \dots, t_n^{(\varepsilon)}\}$. Remarquons que à ε incrément linéaire de tempo tout au long de la séquence correspond α coefficient multiplicatif du tempo initial T_0 qui donne le tempo du dernier événement de la séquence.

Pour cela nous pouvons noter de manière équivalente la séquence $\{t_i^{(\varepsilon)}\} \equiv \{t_i^{(\alpha)}\}$ ².

$$T(t_n^{(\alpha)}) = \alpha T(t_0^{(\alpha)}) = \alpha T_0 = T(t_n^{(\varepsilon)}) \quad (4.1.23)$$

En prenant $\alpha = 2$ par exemple on double le tempo. Notons que $t_0^{(\varepsilon)} = t_0$ car le premier point de la séquence reste fixe indépendamment de l'accélération.

D'après (4.1.18), nous avons :

$$T(t_i^{(\varepsilon)}) = T_0 e^{\varepsilon(t_i^{(\varepsilon)} - t_0)} \quad (4.1.24)$$

(4.1.13) donne :

$$t_n^{(\varepsilon)} - t_0 = (t_n - t_0) \left[\frac{\ln(1 + (t_n - t_0) \cdot \varepsilon)}{(t_n - t_0) \cdot \varepsilon} \right] \quad (4.1.25)$$

Et nous calculons :

$$\alpha = \frac{T(t_n^{(\varepsilon)})}{T_0} = e^{\varepsilon(t_n^{(\varepsilon)} - t_0)} = 1 + (t_n - t_0) \varepsilon \quad (4.1.26)$$

D'où :

$$\varepsilon_\alpha = \frac{\alpha - 1}{t_n - t_0} \quad (4.1.27)$$

ε_α est l'incrément de tempo par pulsation, il donne avec (4.1.18) la courbe de tempo de la séquence $\{t_i\}$ accélérée directement par le coefficient α . La figure **Fig.4.3(a)** présente une séquence rythmique donnée et plusieurs images accélérées de cette séquence, avec α variant de -25% à $+25\%$.

Nous pouvons avec (4.1.27) et (4.1.25) exprimer directement $\{t_i^{(\varepsilon)}\}$ accélération de la séquence $\{t_i\}$ en fonction de α et des t_i uniquement :

$$t^{(\varepsilon)} = t_0 + \frac{t_n - t_0}{\alpha - 1} \ln \left(1 + (\alpha - 1) \frac{t - t_0}{t_n - t_0} \right) \quad (4.1.28)$$

Si l'on désigne par $\{s_i\}$ la *séquence réduite normalisée* de $\{t_i\}$, $\{s_i\}$ commençant à 0 et de longueur 1 :

$$\{s_i\} = \left\{ \frac{t_i - t_0}{t_n - t_0} \right\} \quad (4.1.29)$$

on obtient alors la *séquence réduite non normalisée* accélérée notée $\{s^{(\varepsilon)}\}$ ou de manière

²(α) est un indice de référence à l'accélération uniforme de tempo du début à la fin de la séquence, et non pas un exposant.

équivalente $\{s^{(\alpha)}\}$:

$$s^{(\alpha)} = \frac{\ln(1 + s(\alpha - 1))}{\alpha - 1} \quad (4.1.30)$$

et

$$t^{(\alpha)} = t_0 + (t_n - t_0)s^{(\alpha)} \quad (4.1.31)$$

L'algorithme **Alg.4.1** implémente à partir de (4.1.30) et (4.1.31) l'accélération d'une séquence temporelle quelconque par le coefficient α .

Algorithm 4.1 accélération uniforme directe d'une séquence temporelle

données :

$\{t_0, \dots, t_n\}$, séquence de $n + 1$ points

$\alpha > 0$, coefficient d'accélération

1: **for** $i = 0$ to $n + 1$ **do**

2: séquence réduite normalisée

$$s_i \leftarrow \frac{t_i - t_0}{t_n - t_0}$$

3: accélération directe uniforme

$$s_i^{(\alpha)} \leftarrow \frac{\ln(1 + (\alpha - 1)s_i)}{\alpha - 1} \quad (4.1.32)$$

$$\alpha_i \leftarrow \alpha^{s_i^{(\alpha)}}, \text{ coefficient multiplicatif local du tempo} \quad (4.1.33)$$

4: ré-expansion de la séquence

$$t_i^{(\alpha)} \leftarrow t_0 + (t_n - t_0)s_i^{(\alpha)}$$

5: **end for**

6: **return** $\{t_i^{(\alpha)}\}, \{\alpha_i\}$

La figure **Fig.4.1** présente en parallèle deux accélérations différentes d'une séquence rythmique, mettant en évidence le placement des notes les unes par rapport aux autres dans leurs temporalités propres. Cela serait assez difficile à interpréter musicalement du fait de la complexité créée par la somme des flux d'événements indépendants.

4.1.4 Accélération réciproque uniforme

Nous désignons par accélération réciproque la transformation qui donne une séquence non accélérée c.-à-d. à tempo constant à partir d'une séquence supposée accélérée



FIG. 4.1 – Superposition temporelle d’une séquence rythmique non accélérée – au centre, et accélérée uniformément par les coefficients 0.5 – en haut – et 2 – en bas.

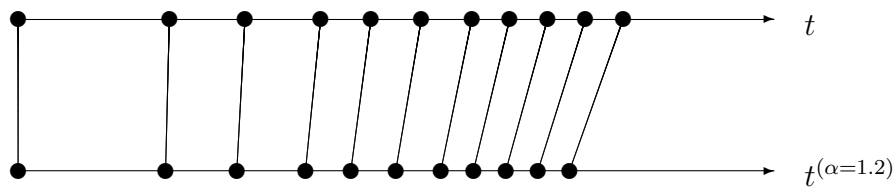


FIG. 4.2 – accélération directe uniforme d’une séquence de 20% , $\alpha = 1.2$

linéairement. Nous ne parlons pas de décélération, la décélération est une accélération par un coefficient inférieur à 1. La fonction réciproque de l’accélération est donc désignée par le terme d’accélération réciproque.

Soit $\{t_0^{(\varepsilon)}, t_1^{(\varepsilon)}, \dots, t_n^{(\varepsilon)}\}$ une séquence donnée – supposée accélérée – de $n + 1$ points. On désire trouver la séquence $\{t_i\}$ correspondante à une accélération réciproque d’un facteur α , soit $\{t_0, t_1, \dots, t_n\}$. $\{t_0, t_1, \dots, t_n\}$ est la réalisation de $\{d_0, d_1, \dots, d_n\}$ à un tempo T_0 constant, i.e. c’est la séquence non accélérée. Or nous supposons une accélération d’un facteur α dans la transcription, soit $\{d_0, d_1, \dots, d_n\}$ avec $T(d_{0_d}) = T_0$ et $T(d_{n_d}) = \alpha T_0$.

Nous utilisons ci-après de manière équivalente la notation (α) ou (ε) utilisée en indice supérieur comme dans (4.1.16) pour désigner la séquence accélérée par opposition à la séquence non accélérée, ainsi on notera de manière équivalente $t_i^{(\varepsilon)}$ ou $t_i^{(\alpha)}$, ε étant l'incrément de tempo dans le domaine symbolique de la partition, et α est le coefficient de multiplication global du tempo du début à la fin de la séquence : le tempo passe de T_0 à αT_0 .

Par définition nous avons (4.1.22), or la réalisation de la séquence notée au tempo T_0 implique $t_n = \frac{d_n}{T_0}$ et :

$$\alpha = \frac{T(d_n)}{T_0} = 1 + \varepsilon(t_n - t_0) \quad (4.1.34)$$

Or d'après l'accélération linéaire réciproque (4.1.16) nous avons :

$$t_n - t_0 = \frac{e^{\varepsilon(t_n^{(\varepsilon)} - t_0)} - 1}{\varepsilon} \quad (4.1.35)$$

Soit

$$\alpha = e^{\varepsilon(t_n^{(\varepsilon)} - t_0)} \quad (4.1.36)$$

Et

$$\varepsilon_\alpha = \frac{\ln(\alpha)}{t_n^{(\varepsilon)} - t_0} \quad (4.1.37)$$

La figure Fig.4.3(b) présente une séquence rythmique accélérée donnée et plusieurs de ses images réciproquement accélérées, avec α variant de -25% à $+25\%$.

Avec la *séquence réduite unitaire* $\{s_i^{(\alpha)}\}$ définie par analogie avec (4.1.29)

$$\{s_i^{(\alpha)}\} = \left\{ \frac{t_i^{(\alpha)} - t_0}{t_n^{(\alpha)} - t_0} \right\} \quad (4.1.38)$$

Nous obtenons en utilisant (4.1.37) dans (4.1.16) :

$$t - t_0 = \frac{e^{\ln \alpha \frac{t^{(\alpha)} - t_0}{t_n^{(\alpha)} - t_0} - 1}}{\frac{\ln \alpha}{t_n^{(\alpha)} - t_0}} \quad (4.1.39)$$

$$t - t_0 = (t_n^{(\alpha)} - t_0) \frac{\alpha^{s^{(\alpha)}} - 1}{\ln \alpha} \quad (4.1.40)$$

$$t - t_0 = (t_n^{(\alpha)} - t_0)s \quad (4.1.41)$$

En posant s la séquence *réduite non unitaire* :

$$s = \frac{\alpha^{s^{(\alpha)}} - 1}{\ln \alpha} \quad (4.1.42)$$

$\{s\}$ est la séquence réciproque de $\{s^{(\alpha)}\}$, son tempo est constant égal à T_0 et,

$$\frac{T(t^{(\alpha)})}{T_0} = \alpha^{s^{(\alpha)}} \quad , \quad t = t_0 + (t_n^{(\alpha)} - t_0)s \quad (4.1.43)$$

Rappelons que $t_0 = t_0^{(\alpha)}$, $\{\alpha^{s^{(\alpha)}}\}$ représente la suite d'accélération locales pour chaque point de la séquence d'origine, t est la séquence – à tempo T_0 constant – réciproque de $t^{(\alpha)}$.

L'algorithme Alg.4.2 implémente à partir de (4.1.43) et (4.1.35) l'accélération uniforme réciproque de coefficient α d'une séquence temporelle quelconque.

La proposition Prop.4.1.1 établit que les fonctions d'accélération et de décélération ainsi définies sont des fonctions linéaires pour toute suite $\{t_n\}$

Proposition 4.1.1. *soit une suite $\{t_n\}$, $\omega \in \mathbb{R}$, $\gamma \in \mathbb{R}^*$ et $\{t'_n\}$ telle que*

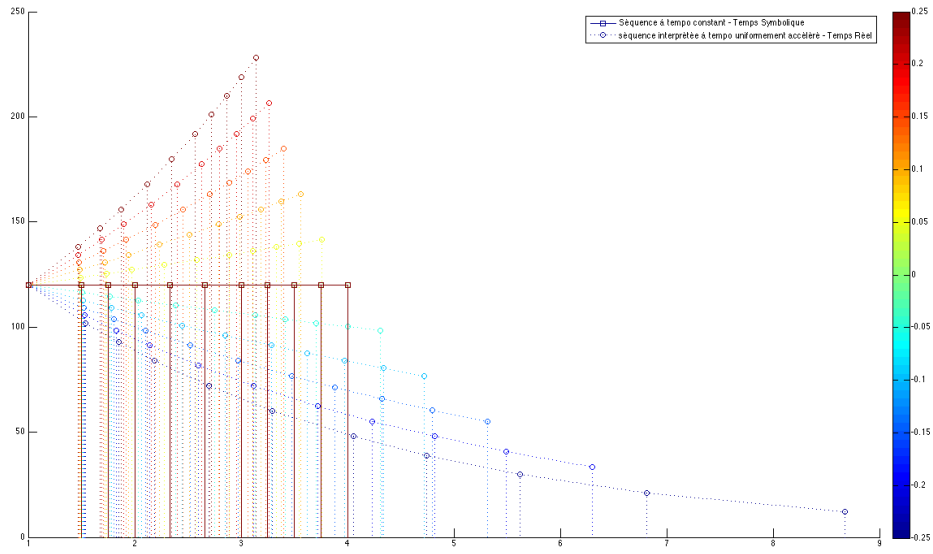
$$\{t'_n\} = \omega + \gamma\{t_n\} \quad (4.1.46)$$

$$\text{alors} \quad acc(\{t'_n\}) = \omega + \gamma acc(\{t_n\}) \quad (4.1.47)$$

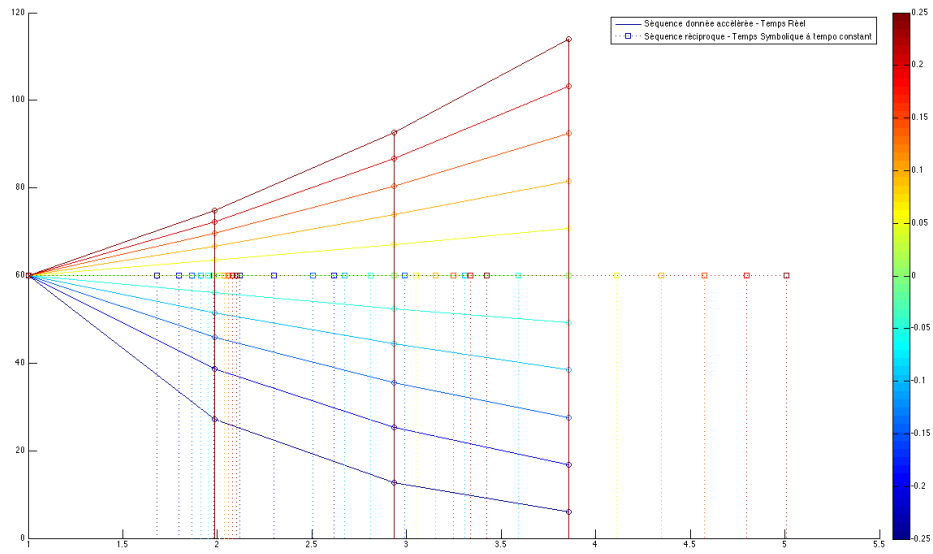
$$\text{et} \quad acc^{-1}(\{t'_n\}) = \omega + \gamma acc^{-1}(\{t_n\}) \quad (4.1.48)$$

démonstration

par calcul direct



(a) Courbe de tempo dans le domaine du Temps Réel, Temps Symbolique vers Temps Réel avec accélération linéaire du tempo



(b) Courbe de tempo réciproque dans le domaine du Temps Symbolique, Temps Réel vers Temps Symbolique, tempo supposé linéairement accéléré

FIG. 4.3 – courbes de tempo

Pour résumer les formules d'accélération continue uniforme directe et réciproque :

$$t^{(\alpha)} = t_0 + \frac{t_n - t_0}{\alpha - 1} \ln \left[1 + \frac{t - t_0}{t_n - t_0} (\alpha - 1) \right] \quad (4.1.49)$$

$$t = t_0 + \frac{t_n^{(\alpha)} - t_0}{\ln \alpha} \left[\alpha \left(\frac{t^{(\alpha)} - t_0}{t_n^{(\alpha)} - t_0} \right) - 1 \right] \quad (4.1.50)$$

Algorithm 4.2 accélération réciproque uniforme**données :** $\{t_0^{(\alpha)}, \dots, t_n^{(\alpha)}\}$, séquence de $n + 1$ points accélérés linéairement de $\alpha > 0$ 1: **for** $i = 0$ to $n + 1$ **do**

2: séquence réduite unitaire

$$s_i^{(\alpha)} \leftarrow \frac{t_i^{(\alpha)} - t_0^{(\alpha)}}{t_n^{(\alpha)} - t_0^{(\alpha)}}$$

3: accélération réciproque uniforme

$$\alpha_i \leftarrow \alpha^{s_i^{(\alpha)}}, \text{ coefficient multiplicatif local du tempo} \quad (4.1.44)$$

$$s_i \leftarrow \frac{\alpha^{s_i^{(\alpha)}} - 1}{\ln \alpha} \quad (4.1.45)$$

4: ré-expansion de la séquence

$$t_i \leftarrow t_0^{(\alpha)} + (t_n^{(\alpha)} - t_0^{(\alpha)})s_i$$

5: **end for**6: **return** $\{t_i\}, \{\alpha_i\}$

4.2 Coïncidence optimale d'une séquence à tempo non constant

Nous sommes maintenant en présence des éléments suivants : nous savons résoudre le problème de coïncidence optimale (3.3.1) et (3.3.2) d'une grille de pulsation à tempo constant pour une séquence temporelle donnée. Nous savons à partir d'une séquence rythmique à tempo supposé accéléré linéairement calculer par inversion la séquence temporelle à tempo constant (4.1.16).

Nous faisons l'hypothèse perceptive que la déformation linéaire du tempo, dans une certaine mesure conserve, à la perception, la même structure de proportionnalité que le rythme joué à tempo constant. C'est à dire qu'à l'écoute, le rythme serait retranscrit de la même manière, qu'il soit joué à tempo constant ou avec une accélération. A notre connaissance, il n'existe pas d'études psycho-cognitives sur cette hypothèse et sur le domaine de validité qu'elle peut avoir – i.e. quelles seraient les limites d'accélération dans lesquelles la perception reste la même ? Ne disposant pas de ces éléments psycho-

cognitifs pour justifier cette hypothèse assez forte, c'est pour le moment une intuition musicale qui nous amène à la poser. Nous voulons alors ramener la séquence de départ $\{t^{(\varepsilon_n)}\}$ dont on ignore si elle est jouée à tempo constant ou non, à la séquence temporelle $\{t_n\}$, son image, à tempo constant, par la relation d'inversion (4.1.16). Nous nous trouvons alors devant un problème d'optimisation de dimension 3 :

$$\min_{s, \Delta, \varepsilon} f_p(s, \Delta, \varepsilon) \quad \Delta \in I, \quad 0 \leq s \leq \Delta, \quad \varepsilon \in E \quad (4.2.1)$$

$$\text{avec } f_p(s, \Delta, \varepsilon) = \sum_{k=1}^n \min_{h=0, \dots, p} |t_k - (s + h\Delta)| \quad (4.2.2)$$

$$\text{et } t_k = t_k^{(\varepsilon)} \left[\frac{e^{\varepsilon t_k^{(\varepsilon)}} - 1}{\varepsilon t_k^{(\varepsilon)}} \right] \text{ avec } \lim_{\varepsilon \rightarrow 0} t_k^{(\varepsilon)} = t_k \quad (4.2.3)$$

$$\text{soit } f_p(s, \Delta, \varepsilon) = \sum_{k=1}^n \min_{h=0, \dots, p} \left| t_k^{(\varepsilon)} \left[\frac{e^{\varepsilon t_k^{(\varepsilon)}} - 1}{\varepsilon t_k^{(\varepsilon)}} \right] - (s + h\Delta) \right| \quad (4.2.4)$$

La complexité mathématique de l'optimisation de (4.2.4) nécessite de trouver une méthode de résolution approchée.

4.2.1 Modèle de grille optimale uniformément accélérée

Nous généralisons maintenant les modèles de grilles optimales précédents – le tempo était constant – afin de pouvoir traiter les cas où le tempo subit une accélération ou une décélération. A cette fin, nous introduisons un facteur d'accélération θ variant dans un intervalle $\mathcal{I}_\theta$. Typiquement pour notre application musicale $\mathcal{I}_\theta = [0.8, 1.2]$, et

- pour $\theta > 1$, le tempo accélère ;
- pour $\theta < 1$, le tempo décélère ;
- pour $\theta = 1$, le tempo est constant.

Etant donné n points de la séquence d'événements, $\{t_1, t_2, \dots, t_n\}$, nous définissons la *séquence réduite unitaire*, ou *séquence normalisée* :

$$s_k = \frac{t_k - t_1}{t_n - t_1} \quad , \quad k = 1, \dots, n \quad (4.2.5)$$

Par définition,

$$0 \leq k \leq 1 \quad (4.2.6)$$

Et notre modèle mathématique – fonction d’erreur avec accélération – devient à l’aide de la relation de réciprocité de l’accélération définie en (4.1.42) :

$$\min_{s, \Delta, \theta} \Phi(s, \Delta, \theta) \quad (4.2.7)$$

Où

$$\Phi(s, \Delta, \theta) = \sum_{k=1}^n \min_{h=0,1,\dots,p} \left| \frac{\theta^{s_k} - 1}{\ln \theta} - s - h\Delta \right| \quad (4.2.8)$$

Il est patique d’introduire, pour toute valeur du paramètre τ , $0 \leq \tau \leq 1$, la fonction :

$$\psi(\theta, \tau) = \frac{\theta^\tau - 1}{\ln \theta}, \quad \theta \in \mathcal{I}_\theta \quad (4.2.9)$$

Quand la valeur de τ est fixée, nous écrivons :

$$\psi(\theta) \equiv \psi(\theta, \tau) \quad (4.2.10)$$

4.2.2 Propriétés de la fonction $\psi(\theta)$

θ étant un coefficient d’accélération du tempo, la fonction $\psi(\theta)$ calcule l’accélération uniforme réciproque d’une séquence. Intuitivement, nous désirons accélérer de façon réciproque une séquence temporelle, c’est-à-dire réaliser une transformation qui va bouger de façon homogène les points accélérés de la séquence donnée vers les points d’une grille régulière à tempo constant. La séquence est présumée accélérée du facteur θ .

Avant d’examiner les propriétés de $\Phi(s, \Delta, \theta)$ (4.2.8) , étudions $\psi(\theta, \tau)$ (4.2.9) (cf. Fig.4.4 et Fig.4.5) telle que :

$$\psi(\theta) = \frac{\theta^\tau - 1}{\ln \theta}, \quad \theta \in \mathcal{I}_\theta, \quad 0 \leq \tau \leq 1 \quad (4.2.11)$$

La séquence normalisée $\{s_k\}$ définie sur l’intervalle $[0, 1]$ a pour image par $\psi(\theta, \tau)$ (4.2.9) une séquence $\{s_k^{(\theta)}\}$ qui repose sur l’intervalle $\left[0, \frac{\theta - 1}{\ln \theta}\right]$. Fig.4.4 montre le délacement continu des points du segment $[0, 1]$ soumis à différentes valeurs fixées

pour l'accélération réciproque θ . On voit sur **Fig.4.4** comment une pulsation régulière jouée avec un *accelerando* continu devient une pulsation régulière sans *accelerando* après accélération réciproque.

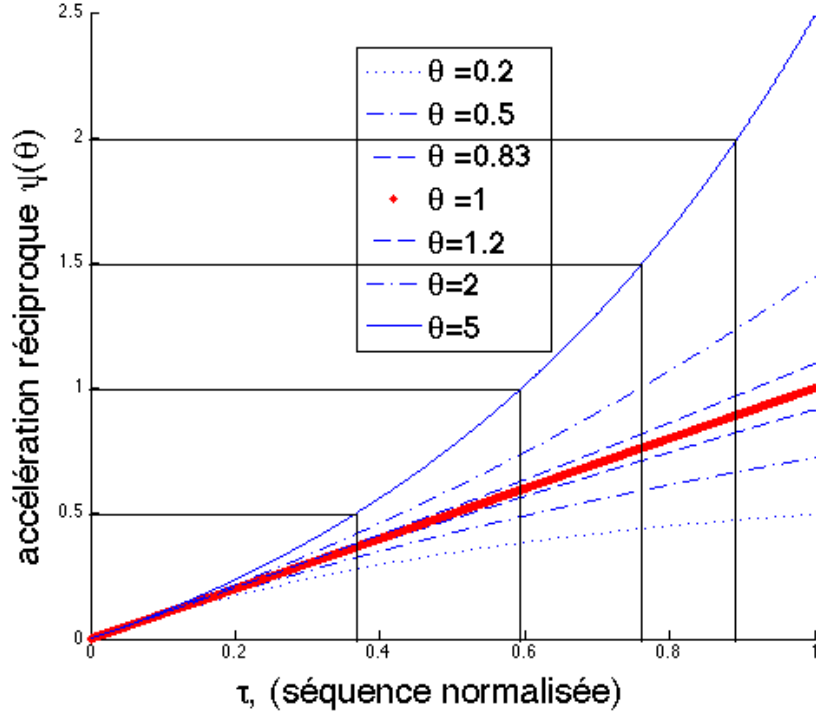


FIG. 4.4 – Accélération réciproque $\psi(\theta, \tau)$

Proposition 4.2.1. *On a, pour $\theta \in \mathcal{I}_\theta$*

$$\psi'(\theta) = \frac{\tau\theta^{\tau-1}}{\ln \theta} - \frac{\theta^\tau - 1}{\theta \ln^2 \theta} \quad (4.2.12)$$

$$\psi''(\theta) = \frac{2(\theta^\tau - 1)}{\theta^2 \ln^3 \theta} + \frac{\theta^\tau - 1}{\theta^2 \ln^2 \theta} - \frac{2\tau\theta^{\tau-2}}{\ln^2 \theta} + \frac{(\tau - 1)\tau\theta^{\tau-2}}{\ln \theta} \quad (4.2.13)$$

Proposition 4.2.2. *Pour $\theta = 1$, la fonction $\psi(\theta)$ est indéfinie. Néanmoins,*

$$\lim_{\theta \rightarrow 1} \psi(\theta) = \tau \quad (4.2.14)$$

Ainsi, si on définit $\psi(1) = \tau$, la fonction $\psi(\theta)$ est continue à $\theta = 1$.

Démonstration

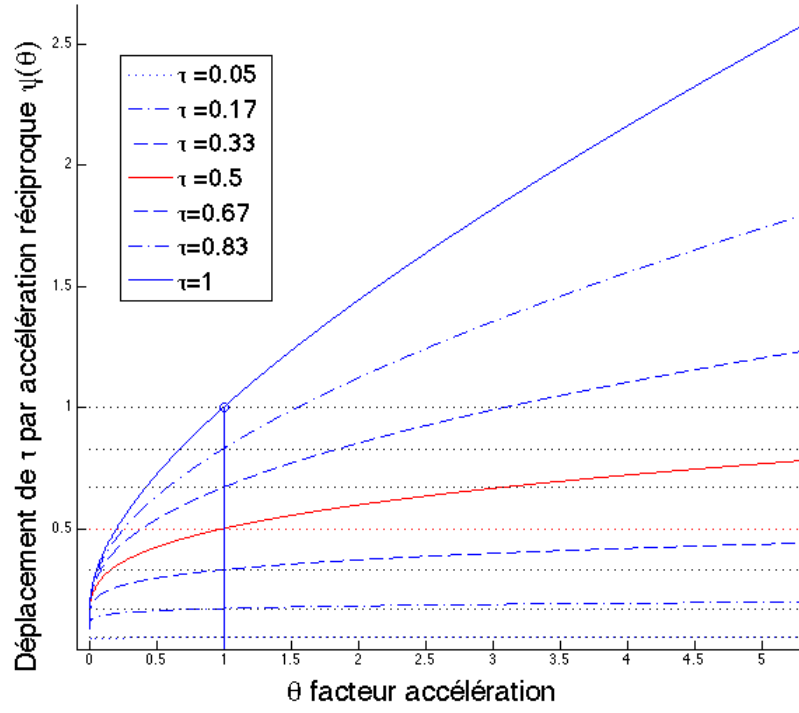
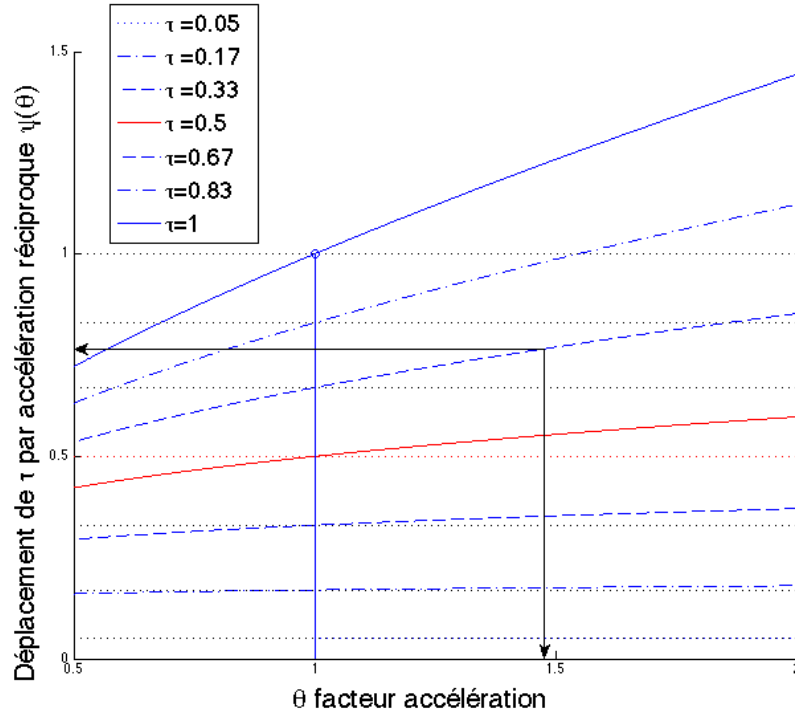
(a) $\theta \in [0.1, 10]$ (b) $\theta \in [0.5, 2]$

FIG. 4.5 – Accélération réciproque $\psi(\theta, \tau)$. Sur chaque courbe à $\tau = s_k$ constant, nous lisons l'image de la séquence réduite s_k accélérée réciproquement du facteur θ . Nous pouvons alors trouver graphiquement la solution θ_{kh} de ξ_{kh} (4.2.38) non inversible analytiquement

$$\lim_{\theta \rightarrow 1} \psi(\theta) = \lim_{\theta \rightarrow 1} \frac{\theta^\tau - 1}{\ln \theta} = \lim_{\theta \rightarrow 1} \frac{\tau \theta^{\tau-1}}{\frac{1}{\theta}} = \tau \lim_{\theta \rightarrow 1} \theta^\tau = \tau$$

Proposition 4.2.3.

$$\psi(\theta, 0) = 0 \quad \forall \theta \in \mathcal{I}_\theta \quad (4.2.15)$$

$$\psi(\theta, 1) = \frac{\theta - 1}{\ln \theta} \quad \forall \theta \in \mathcal{I}_\theta \quad (4.2.16)$$

Proposition 4.2.4. *Pour $\theta = 1$, les fonctions $\psi'(\theta)$, $\psi''(\theta)$ sont indéfinies, néanmoins*

$$\lim_{\theta \rightarrow 1} \psi'(\theta) = \frac{\tau^2}{2} \quad (4.2.17)$$

$$\lim_{\theta \rightarrow 1} \psi''(\theta) = \frac{\tau^2}{6}(2\tau - 3) \quad (4.2.18)$$

Ainsi, si on définit $\psi'(1) = \frac{\tau^2}{2}$ et $\psi''(1) = \frac{\tau^2}{6}(2\tau - 3)$, les fonctions $\psi'(\theta)$ et $\psi''(\theta)$ sont continues en $\theta = 1$.

Théorème 4.2.5. *La fonction $\psi(\theta)$ définie sur $]0 + \infty[$ est concave.*

Démonstration

Nous montrons pour cela que $\psi'(\theta)$ est positive et $\psi''(\theta)$ est négative.

$\psi'(\theta)$ est continue dérivable sur $]0 + \infty[$:

$$\psi'(\theta) = \frac{\tau \theta^{\tau-1}}{\ln \theta} - \frac{\theta^\tau - 1}{\theta \ln^2 \theta} \quad (4.2.19)$$

$$\text{avec } \psi'(1) = \frac{\tau^2}{2} > 0 \quad (4.2.20)$$

Montrons que $\psi'(\theta) > 0$ pour $\theta \neq 1$, soit :

$$\tau \theta^\tau \ln \theta > \theta^\tau - 1 \quad (4.2.21)$$

$$1 > \theta^\tau (1 - \tau \ln \theta) \quad (4.2.22)$$

$$\theta^{-\tau} = \frac{1}{\theta^\tau} > 1 + \ln \theta^{-\tau} \quad (4.2.23)$$

$$\text{Soit en posant } \zeta = \theta^{-\tau} > 0, \quad 1 + \ln \zeta < \zeta \quad (4.2.24)$$

(4.2.24) est vrai et est un résultat connu : $\zeta - 1$ est la droite tangente au point $\zeta = 1$

de la courbe $\ln \zeta$ qui se situe sous sa tangente en ce point, i.e. $\ln \zeta < \zeta - 1$ est vérifié. En conséquence, avec (4.2.20) $\psi'(\theta)$ est positive sur $]0 + \infty[$.

$\psi''(\theta)$ est continue dérivable sur $]0 + \infty[$ et nous avons :

$$\psi''(\theta) = \frac{2(\theta^\tau - 1)}{\theta^2 \ln^3 \theta} + \frac{\theta^\tau - 1}{\theta^2 \ln^2 \theta} - \frac{2\tau\theta^{\tau-2}}{\ln^2 \theta} + \frac{(\tau - 1)\tau\theta^{\tau-2}}{\ln \theta} \quad (4.2.25)$$

$$\text{Avec } \psi''(1) = \frac{\tau^2}{6}(2\tau - 3) < 0 \quad (4.2.26)$$

Montrons que $\psi''(\theta) < 0$ pour $\theta \neq 1$.

(4.2.25) donne :

$$\psi''(\theta) = \frac{2(\theta^\tau - 1) + ((1 - 2\tau)\theta^\tau - 1)\ln \theta + (\tau - 1)\tau\theta^\tau \ln^2 \theta}{\theta^2 \ln^3 \theta} \quad (4.2.27)$$

$$\text{Que nous notons : } \psi''(\theta) = \frac{\psi''_{Num}(\theta)}{\psi''_{Den}(\theta)} \quad (4.2.28)$$

$$\psi''_{Den}(\theta) \text{ de (4.2.28) est négatif sur }]0, 1[\text{ et positif sur }]1, \infty[. \quad (4.2.29)$$

Regardons alors le numérateur $\psi''_{Num}(\theta)$ de (4.2.28). En prenant sa dérivée, il vient :

$$\begin{aligned} \psi'''_{Num}(\theta) &= 2\tau\theta^{\tau-1} + \frac{(1 - 2\tau)\theta^\tau - 1}{\theta} + (1 - 2\tau)\tau\theta^{\tau-1} \ln \theta \\ &\quad + 2(\tau - 1)\tau\theta^{\tau-1} \ln \theta + (\tau - 1)\tau^2\theta^{\tau-1} \ln^2 \theta \end{aligned} \quad (4.2.30)$$

$$= \frac{\theta^\tau - 1 - \tau\theta^\tau \ln \theta + (\tau - 1)\tau^2\theta^\tau \ln^2 \theta}{\theta} \quad (4.2.31)$$

Le dernier terme de (4.2.31) est négatif car $0 \leq \tau \leq 1$ et

$$\theta^\tau - 1 - \tau\theta^\tau \ln \theta < 0 \quad (4.2.32)$$

$$\text{car } 1 - \theta^{-\tau} + \ln \theta^{-\tau} < 0 \quad (4.2.33)$$

$$\text{Soit en posant } \zeta = \theta^{-\tau} > 0 \quad \ln \zeta < \zeta - 1 \quad (4.2.34)$$

(4.2.34) implique (4.2.32) négatif et donc $\psi'''_{Num}(\theta)$ (4.2.31) négatif sur $\theta \in]0 + \infty[$. Ainsi $\psi''_{Num}(\theta)$ décroît sur $\theta \in]0 + \infty[$, or $\psi''_{Num}(1) = 0 \implies \psi''_{Num}(\theta)$ positif sur $]0, 1[$ et négatif sur $]1 + \infty[$.

$\psi''(\theta)$ étant continue dérivable sur $\theta \in]0 + \infty[$, avec (4.2.29) nous obtenons $\psi''(\theta)$ négatif sur $]01[$ et négatif sur $]1 + \infty[$, soit finalement avec (4.2.26) $\psi''(\theta)$ négatif sur $]0 + \infty[$

□

Nous utiliserons le Th.4.2.5 dans Alg.4.4 pour la recherche des racines de (ξ_{hk}) (4.2.38) à l'aide de la méthode de Newton.

4.2.3 Propriétés de la fonction $\Phi(s, \Delta, \theta)$

$\Phi(s, \Delta, \theta)$ (4.2.8) est la fonction que nous désirons minimiser suivant ses trois paramètres afin de trouver les valeurs les plus probables de Δ , s et θ . Cela revient à replacer la séquence temporelle donnée dans un tempo constant pour lequel nous calculons la *grille optimale* développée précédemment.

Pour s et Δ donnés, posons

$$\Psi(\theta) \equiv \Phi(s, \Delta, \theta) \quad , \quad \theta \in \mathcal{I}_\theta \quad (4.2.35)$$

avec $\Phi(s, \Delta, \theta)$ définie en (4.2.8). On peut écrire de manière plus compacte :

$$\Psi(\theta) = \sum_{k=1}^n \min_{h=0,1,\dots,p} |\psi(\theta, s_k) - g_h| \quad (4.2.36)$$

où nous avons

$$g_h = s + h\Delta, \quad h = 0, 1, \dots, p \quad (4.2.37)$$

Pour chacun des couples d'index (k, h) , considérons l'équation :

$$(\xi_{hk}) \quad \frac{\theta^{s_k} - 1}{\ln \theta} = s + h\Delta \quad (4.2.38)$$

ou équivalent

$$(\xi_{hk}) \quad \psi(\theta, s_k) = g_h \quad (4.2.39)$$

Soit θ_{hk} la solution unique – si elle existe – de (ξ_{hk}) (4.2.38) dans l'intervalle $\mathcal{I}_\theta$. Une valeur $\theta \in \mathcal{I}_\theta$ sera une racine de Ψ si θ est une racine d'un (ξ_{hk}) .

Soit $\theta \in \mathcal{I}_\theta$, nous nous intéressons aux propriétés suivantes de θ par rapport à Ψ :

1. θ est une racine de Ψ
2. θ est un point anguleux de Ψ (i.e. Ψ n'est pas dérivable en θ)
3. θ est un minimum local de Ψ

Proposition 4.2.6. *Si θ est une racine, alors θ est un point anguleux*

Conjecture 4.2.7. *θ est un minimum local si et seulement si θ est une racine*

4.3 Algorithme de minimisation de $\Phi(s, \Delta, \theta)$

Nous décrivons ci-dessous un algorithme de minimisation du problème (4.2.7). Nous ne savons pas encore si notre algorithme trouve ou non un minimum global de $\Phi(s, \Delta, \theta)$ sur tous les (s, Δ, θ) possibles.

4.3.1 un schéma itératif alterné

Nous envisageons le schéma itératif suivant pour le problème (4.2.7). Nous commençons l'algorithme avec une estimation $(\hat{s}, \hat{\Delta}, \hat{\theta})$. Dénotons par $(\hat{s}, \hat{\Delta}, \hat{\theta})$ la solution incombante – i.e. la meilleur solution trouvée jusqu'à présent – au début d'une itération donnée. L'itération consiste en deux étapes :

1. minimiser $\Phi(s, \Delta, \hat{\theta})$ sur s et Δ . Posons $(\bar{s}, \bar{\Delta})$ le minimum – global – de $\Phi(s, \Delta, \hat{\theta})$
2. minimiser $\Phi(\bar{s}, \bar{\Delta}, \theta)$ sur tous les $\theta \in \mathcal{I}_\theta$. Posons $\bar{\theta}$ le minimum de $\Phi(\bar{s}, \bar{\Delta}, \theta)$

L'algorithme s'arrête si $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) = \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$, i.e. si l'itération est incapable d'améliorer la valeur incombante $\Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$ – en ce sens l'algorithme est descendant. Sinon – i.e. si $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) < \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$ le triplet $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta})$ devient la nouvelle solution incombante $(\hat{s}, \hat{\Delta}, \hat{\theta})$ et l'itération recommence. L'Alg.4.3 décrit formellement cette procédure.

Remarques

1. Si la condition *until* n'est pas satisfaite, alors on doit avoir $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) < \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$ et donc $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta})$ peut être considérée comme la nouvelle solution incombante

Algorithm 4.3 descente alternative pour la minimisation de $\Phi(s, \Delta, \theta)$

données :

$\{t_1, \dots, t_n\}$, $(t_1 < \dots < t_n)$ séquence de n points
 $\theta^{(0)} > 0$, choix à priori sur le coefficient d'accélération

1: *initialisation* – les valeurs de $\bar{s}$ et $\bar{\Delta}$ peuvent être arbitraires

$$\bar{\theta} \leftarrow \theta^{(0)}$$

$$\bar{s} \leftarrow 0$$

$$\bar{\Delta} \leftarrow 1$$

2: **repeat**

$$3: (\hat{s}, \hat{\Delta}, \hat{\theta}) \leftarrow (\bar{s}, \bar{\Delta}, \bar{\theta})$$

$$4: \Phi(\bar{s}, \bar{\Delta}, \hat{\theta}) \leftarrow \min_{(s, \Delta)} \Phi(s, \Delta, \hat{\theta})$$

$$5: \Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) \leftarrow \min_{\theta \in \mathcal{I}_\theta} \Phi(\bar{s}, \bar{\Delta}, \theta)$$

$$6: \textbf{until } \Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) = \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$$

$$7: \textbf{return } (\hat{s}, \hat{\Delta}, \hat{\theta})$$

2. En pratique, la condition d'arrêt $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) = \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$ peut être remplacée par

$$\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) \geq \Phi(\hat{s}, \hat{\Delta}, \hat{\theta}) - \varepsilon$$

où ε est une tolérance décidée par l'utilisateur.

4.3.2 Première étape de minimisation

Afin de trouver pour un $\hat{\theta}$ donné,

$$\Phi(\bar{s}, \bar{\Delta}, \hat{\theta}) = \min_{(s, \Delta)} \Phi(s, \Delta, \hat{\theta})$$

On procède de la façon suivante, posons :

$$\hat{t}_k = \frac{\hat{\theta}^{s_k} - 1}{\ln \hat{\theta}}, \quad k = 1, \dots, n$$

puis

$$\Phi(s, \Delta, \theta) = \sum_{k=1}^n \min_{h=0,1,\dots,p} |\hat{t}_k - s - h\Delta| \quad (4.3.1)$$

Nous pouvons trouver $\min_{h=0,1,\dots,p} |\hat{t}_k - s - h\Delta|$ par les méthodes discutées précédemment (Alg.3.6).

4.3.3 Seconde étape de minimisation

Afin de trouver pour un $(\bar{s}, \bar{\Delta})$ donné,

$$\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) = \min_{\theta \in \mathcal{I}_\theta} \Phi(\bar{s}, \bar{\Delta}, \theta)$$

Nous utilisons la propriété de la conjecture (4.2.7) qui spécifie que les minimum locaux de Φ sont précisément les racines de Φ . Soit, comme précédemment, θ_{hk} l'unique racine de (ξ_{hk}) (4.2.38) – ou son équivalent (4.2.39). Alors

$$\min_{\theta \in \mathcal{I}_\theta} \Phi(\bar{s}, \bar{\Delta}, \theta) = \min_{(h,k)} \Phi(\bar{s}, \bar{\Delta}, \theta_{hk}) \quad , \quad h = 0, \dots, p \text{ et } k = 1, \dots, n \quad (4.3.2)$$

Nous discrétisons ainsi l'intervalle $\mathcal{I}_\theta$ par $n(p+1)$ racines θ_{hk} . Nous pouvons calculer θ_{hk} par la méthode de Newton dont la convergence superlinéaire [Dontchev et al., 1999] [Levin et al., 2000] est assurée par le Th.4.2.5. La procédure est décrite par Alg.4.4

Algorithm 4.4 algorithme de Newton pour trouver la racine θ_{hk} de (ξ_{hk}) (4.2.39)

données :

$h \in \{0, 1, \dots, p\}$
 $k \in \{0, 1, \dots, n\}$
 $0 < s_k \leq 1$
 $\bar{\theta}_{hk}$, valeur à priori de θ_{hk}
 $(\bar{s}, \bar{\Delta})$, centre et pas de grille
 ε , approximation désirée pour la racine

```

1:  $\bar{\theta} \leftarrow \bar{\theta}_{hk}$ 
2: repeat
3:    $\hat{\theta} \leftarrow \bar{\theta}$ 
4:    $\bar{\theta} \leftarrow \hat{\theta} - \frac{\psi(\hat{\theta}, s_k)}{\psi'(\hat{\theta}, s_k)}$ 
5: until  $|\psi(\hat{\theta}, s_k) - g_h| \leq \varepsilon$ 
6:  $\theta_{hk} \leftarrow \bar{\theta}$ 
7: return  $\theta_{hk}$ , racine de  $(\xi_{hk})$ 

```

Dans la description de Alg.4.4, $\psi(\theta, \tau)$, $\psi'(\theta, \tau)$ et g_h sont donnés respectivement par

(4.2.9), (4.2.12) et (4.2.37). Pour chaque paire (h, k) , l'algorithme de Newton converge très rapidement vers une racine θ_{hk} de l'équation (ξ_{hk}) (4.2.38), à condition que le choix initial $\bar{\theta}_{hk}$ soit assez proche de θ_{hk} .

Une approche échantillonnée pour trouver une bonne valeur initiale

Soit N un entier positif assez grand - e.g. $N = 100$. Divisons alors $\mathcal{I}_\theta$ en N sous-intervalles $[\theta_{i-1}, \theta_i], i = 1, \dots, N$ de longueur égale $\frac{\text{longueur}(\mathcal{I}_\theta)}{N}$

Définissons pour chaque couple (h, k)

$$\bar{\theta}_{hk} = \theta_{i^*}$$

où $|\psi(\theta_{i^*}, s_k) - g_h| = \min_{i=0,1,\dots,N} |\psi(\theta_i, s_k) - g_h|$

Alors $\bar{\theta}_{hk}$ est un choix initial raisonnablement proche de θ_{hk} .

Une autre méthode pour trouver une racine de (ξ_{hk}) (4.2.38)

On sait que, quand x est proche de 0

$$\ln(1+x) \sim x \quad (4.3.3)$$

Effectuons alors le changement de variable

$$\eta = \theta - 1 \quad (4.3.4)$$

(ξ_{hk}) (4.2.38) devient alors avec (4.3.4)

$$\frac{(1+\eta)^{s_k} - 1}{\eta} \sim g_h \quad (4.3.5)$$

soit

$$\eta \sim \frac{(1+\eta)^{s_k} - 1}{g_h} \quad (4.3.6)$$

La réaction d'approximation (4.3.6) suggère le schéma itératif suivant pour trouver une solution de l'équation :

$$\eta = \frac{(1+\eta)^{s_k} - 1}{g_h} \quad (4.3.7)$$

i.e. un point fixe de la correspondance

$$\eta \mapsto \frac{(1 + \eta)^{s_k} - 1}{g_h} \quad (4.3.8)$$

Le schéma itératif est alors

$$\begin{aligned} \eta^{(0)} & \text{ (arbitraire)} \\ \eta^{(i)} &= \frac{(1 + \eta^{(i-1)})^{s_k} - 1}{g_h}, \quad i = 1, 2, \dots \end{aligned} \quad (4.3.9)$$

Nous ne savons néanmoins pas si

1. la séquence $\{\eta^{(i)}\}_{i=0,1,2,\dots}$ converge
2. si oui, alors si la limite coïncide avec la racine θ_{hk}

De toute façon, si les deux remarques précédentes ne sont plus vérifiées, nous pouvons toujours arrêter l'algorithme après un nombre n d'itérations suffisant, et utiliser $\eta^{(m)}$ comme choix initial de $\bar{\theta}_k$ dans l'algorithme de Newton.

4.3.4 Algorithme généralisé pour minimiser $\Phi(s, \Delta, \theta)$

Nous sommes maintenant en mesure de proposer un algorithme complet **Alg.4.5** pour la minimisation de $\Phi(s, \Delta, \theta)$ (4.2.7) intégrant la méthode de Newton pour la seconde étape de minimisation.

La figure **Fig.4.6** illustre le fonctionnement de l'algorithme **Alg.4.5** sur une séquence de points accélérée. À chaque itération, on procède par accélération réciproque uniforme : le meilleur facteur d'accélération θ est recherché parmi les valeurs discrètes possibles mises en évidence par la résolution des (ξ_{hk}) . Si aucun θ_{hk} ne permet de trouver une grille optimale meilleure que celle de l'itération précédente, alors l'algorithme s'arrête. En pratique, la première itération est déjà suffisante pour une bonne approximation de θ . La convergence de l'algorithme n'est pas encore mathématiquement prouvée il faudrait démontrer que $\bar{\theta}$ est toujours une racine de (ξ_{hk}) ce. Si tel est le cas, alors il s'agit d'un nouveau théorème. La difficulté de la démonstration réside dans la résolution non inversible qui utilise la méthode de Newton.

Algorithm 4.5 minimisation de $\Phi(s, \Delta, \theta)$

données : $\{t_1, \dots, t_n\}$, ($t_1 < \dots < t_n$) séquence de n points,
 p nombre de points de grille

1: *Initialisation*

$$\begin{aligned} \{s_k\} &= \frac{\{t_k\} - t_1}{t_n - t_1} \quad , \quad k = 1, \dots, n \\ \bar{s} &\leftarrow 0 \\ \bar{\Delta} &\leftarrow 1 \\ \bar{\theta} &\leftarrow \theta^{(0)} \leftarrow 1 \end{aligned}$$

2: **repeat**

3:
$$(\hat{s}, \hat{\Delta}, \hat{\theta}) \leftarrow (\bar{s}, \bar{\Delta}, \bar{\theta}) \quad (4.3.10)$$

4: *Première étape*

$$\Phi(\bar{s}, \bar{\Delta}, \hat{\theta}) \leftarrow \min_{(s, \Delta)} \Phi(s, \Delta, \hat{\theta}), \quad (\text{avec Alg. 3.6}) \quad (4.3.11)$$

5: *Seconde étape*

6: **for** $k = 1$ to n **do**

7: **for** $h = 0$ to p **do**

8:
$$\bar{\theta}_{hk} \leftarrow \hat{\theta} \quad (4.3.12)$$

9: **repeat**

10:
$$\hat{\theta}_{hk} \leftarrow \bar{\theta}_{hk} \quad (4.3.13)$$

$$\bar{\theta}_{hk} \leftarrow \hat{\theta}_{hk} - \frac{\psi(\hat{\theta}_{hk}, s_k) - g_h}{\psi'(\hat{\theta}_{hk}, s_k)} \quad (4.3.14)$$

11: **until** $|\psi(\hat{\theta}_{hk}, s_k) - g_h| \leq \varepsilon$

12:
$$\theta_{hk} \leftarrow \bar{\theta}_{hk}, \quad \text{racine de } (\xi_{hk}) \quad (4.3.15)$$

13: **end for**

14: **end for**

15:
$$\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) \leftarrow \min_{\theta \in \{\theta_{hk}\}} \Phi(\bar{s}, \bar{\Delta}, \theta) \quad (4.3.16)$$

$$\sigma(\bar{s}, \bar{\Delta}, \bar{\theta}) \leftarrow \frac{2\Phi(\bar{s}, \bar{\Delta}, \bar{\theta})}{n\bar{\Delta}}, \quad \text{indice de coïncidence} \quad (4.3.17)$$

16: **until** $\Phi(\bar{s}, \bar{\Delta}, \bar{\theta}) = \Phi(\hat{s}, \hat{\Delta}, \hat{\theta})$

17: **return** $(\hat{s}, \hat{\Delta}, \hat{\theta}), \sigma(\bar{s}, \bar{\Delta}, \bar{\theta})$



FIG. 4.6 – algorithme itératif Alg.4.5 sur une séquence supposée accélérée, on remarque que dès la première itération de θ la suite de point est proche de la suite limite de convergence

4.4 Utilisation musicale de la grille optimale

La grille optimale nous permet de proposer un tempo pour une séquence d'événements temporels. Mais comment cette relation mathématique permet effectivement ce choix en situation musicale et quelles en sont les implications ?

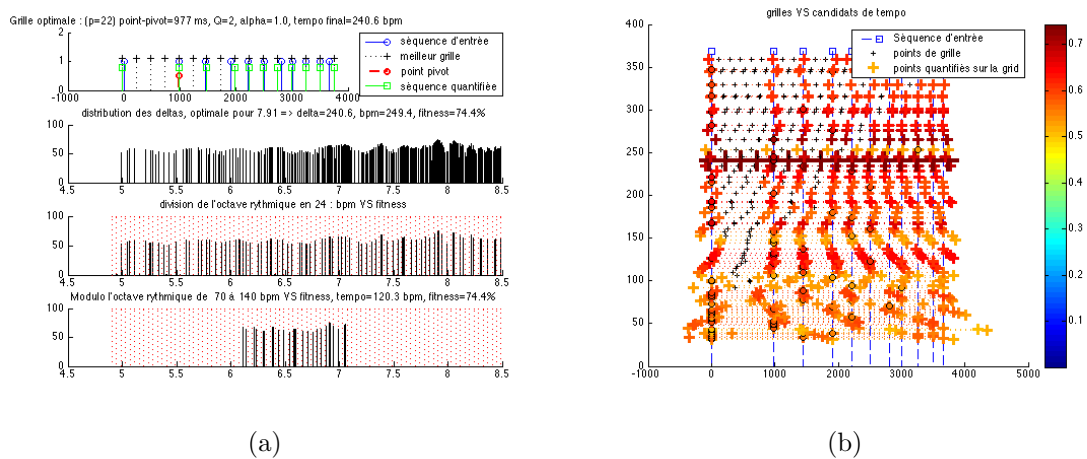


FIG. 4.7 – Grille Optimale pour les points de la séquence supposée non accélérée $\{0, 0.97, 1.44, 1.90, 2.20, 2.49, 2.79, 2.99, 3.24, 3.47, 3.64\}$

Les figures Fig.4.7, Fig.4.8 et Fig.4.9 montrent à travers les possibilités de grille optimale, les probabilités de pulsation pour une séquence de points considérée non accélérée ou accélérée d'un facteur $\alpha = 1.1$ en Fig.4.8 et $\alpha = 1.2$ en Fig.4.9.

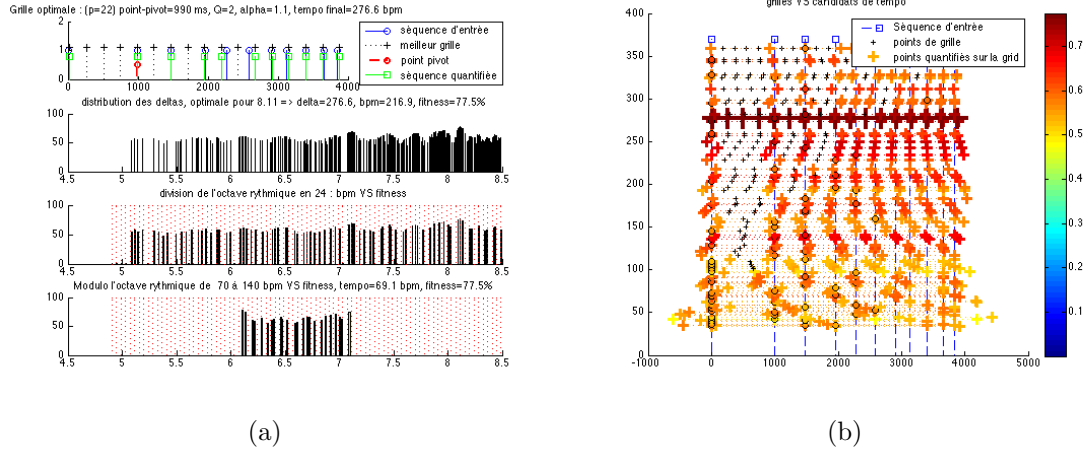


FIG. 4.8 – Grille Optimale, séquence supposée accélérée de $\alpha = 1.1$, les points de la séquence accélérée réciproque résultante sont $\{0, 0.98, 1.47, 1.95, 2.26, 2.58, 2.89, 3.11, 3.38, 3.64, 3.82\}$

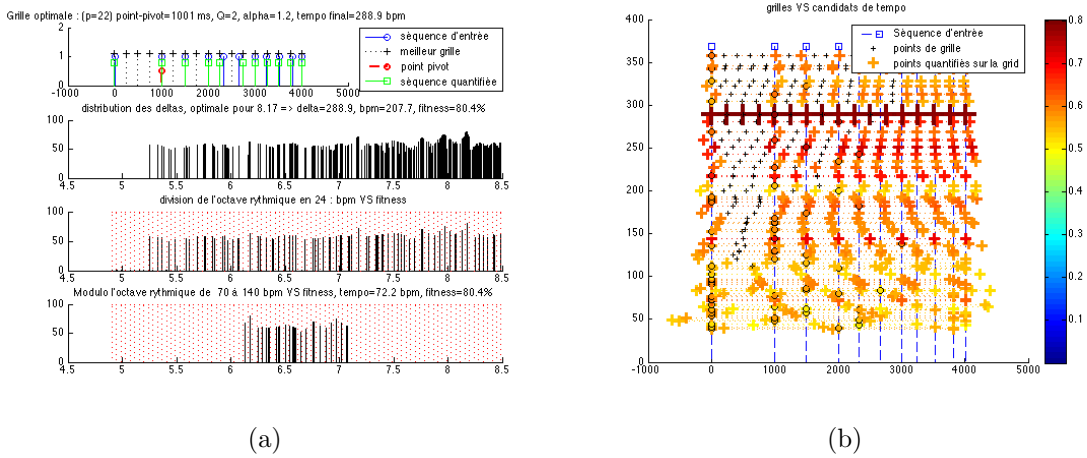


FIG. 4.9 – Grille Optimale, séquence supposée accélérée de $\alpha = 1.2$, les points de la séquence accélérée réciproque résultante sont $\{0, 1, 1.5, 2, 2.33, 2.66, 3, 3.24, 3.53, 3.81, 4.01\}$

4.4.1 Transcription rythmique et modèle probabiliste

Nous avons vu précédemment que le tempo s'exprime dans deux domaines temporels. Le domaine symbolique représente le temps sous forme d'unités ou de pulsations, il est

comme une partition des événements. Le domaine du temps réel est le temps qui porte les instants de la réalisation, de la réalité que nous percevons. On peut considérer cette transformation du symbolique au réel ou du réel vers le symbolique comme une *distorsion dynamique relative* : les deux temporalités glissent l'une par rapport à l'autre. Dans ce contexte, la transcription rythmique est alors une distorsion particulière qui résulte d'un choix de valeurs rythmiques et d'une courbe de tempo. On comprends alors qu'il n'y a pas une transcription rythmique mais plusieurs transcriptions rythmiques possibles associées à une réalisation d'événements. L'exemple très simple qui suit le montre : on peut percevoir une note qui dure une seconde comme une noire à 60 ou bien une blanche à 120, ce sont deux transcriptions rythmiques possibles mais l'une est certainement plus probable que l'autre suivant le contexte musical. Maintenant imaginons que ce contexte change dynamiquement et que nous soyons à chaque moment dans cette recherche de transcription articulé par la courbe de tempo qui est en chacun de ses points le passage d'un domaine à un autre. La transcription à l'aide d'une grille optimale de coïncidence permet dans une certaine mesure de s'affranchir des variations dues à l'interprétation. En fixant une résolution de quantification, on considère transcription et quantification dans une même opération. Le choix de la transcription est alors la question difficile à résoudre, et le choix du tempo en dépend également.

Nous pouvons maintenant répondre à la question concernant le lien entre le tempo et la quantification : le tempo ne résulte pas de la quantification, ni celle-ci du tempo, même si leur relation est évidente. C'est de la transcription rythmique, qui est finalement un codage inféré à partir de critères d'optimisation, que résultent le suivi de tempo et la quantification associée.

Il s'opère ici un choix dynamique qui maintient la relation de tempo et de quantification. Intuitivement nous désirons utiliser un critère de parcimonie pour lequel nous obtenons une notation relativement simple associée à une courbe de tempo qui varie le moins possible ou le mieux possible (un changement brusque de tempo serait par contre différent d'une accélération de tempo continue). Autrement dit, nous cherchons dynamiquement la notation la plus simple pour la variation de la courbe de tempo la plus simple afin d'obtenir une transcription et un suivi de tempo associé.

[Cemgil, 2004] voit le problème de la transcription comme un problème d'estimation d'états pour lequel il faut identifier la trajectoire la plus probable. Il s'agit de considérer les quantités que nous devons inférer comme une séquence de variables cachées notées X . Nous définissons alors un modèle probabiliste qui relie les observations Y aux X cachées avec éventuellement un ensemble de paramètres contextuels fixes Θ . Etant

données les observations, la transcription peut être vue comme un problème général d'inférence bayésienne où nous calculons une distribution postérieure – ou *posterior* – sur des quantités cachées en inversant le modèle à l'aide du théorème de Bayes. Le chapitre suivant propose un modèle d'inférences probabilistiques, bayésien ou markovien au sens général.

Chapitre 5

Suivi de tempo

Nous prolongeons dans un cadre dynamique les résultats obtenus au chapitre (3) avec l'algorithme de *grille optimale accélérée* Alg.4.5 qui généralise l'algorithme de *grille optimale* Alg.3.6. Les développements mathématiques du chapitre (3) donnent une estimation locale du tempo sur un ensemble de n points avec une prise en compte de l'accélération locale. Dans une situation temporelle dynamique, cette fenêtre de n points représente la mémoire à court terme des n derniers événements. Afin de comprendre nos hypothèses, nous les mettons en relation avec certaines recherches physiologiques et de perception psycho-cognitives. Les mécanismes sous-jacents de la perception du tempo ont été étudiés récemment à l'aide de corrélations de données provenant de l'électrophysiologie du cerveau [Pfeuty et al., 2003]. Ces résultats mettent à jour que des processus physiologiques basés sur les intervalles de temps interviennent très probablement dans l'élaboration de la trace de la mémoire des intervalles. Sans évoquer *a priori* un contexte psycho cognitif, l'étude théorique menée tout au long du chapitre (3) sur une suite d'événements, qui est également une suite d'intervalles temporels, s'appuie sur la génération d'une série finie dénombrable de sous-intervalles. Cette série est engendrée par la série d'origine en divisant chaque intervalle entre deux événements par un nombre entier donné p que nous avons démontré au Th.3.3.2 et Alg.3.6 équivalent au nombre total de points de la grille optimale qui relie les événements. p représente la résolution minimale de quantification et décrit le nombre maximal de points de la grille optimale.

La grille sous-jacente des pulsations ou *tactus* est un compromis entre la notation ou transcription la plus facile et la meilleure variation du tempo. La transcription dynamique, ou le décodage du suivi de tempo, s'effectue à partir de l'estimation de tempo qui se propage dynamiquement. Nous avons vu en (4.4) qu'il n'existe pas une seule solution de transcription, mais plusieurs et que chacune se rapporte à un degré de certitude ou de probabilité. La grille optimale agit en tant que quantificatrice des événements qui la créent. Quant à la perception rythmique qui en résulte, les travaux de [Pfeuty et al., 2003] et de [Chen et al., 2008] éclairent notre démarche et pourraient établir qu'il existe probablement une forme d'auto organisation des événements par notre système cognitif qui cherche à établir des symétries pour reconnaître et discerner une structure organisatrice temporelle basée sur les *inter-onsets* et un sous-ensemble généré de leurs sous-multiples. La mémoire à cours terme et la mémoire rétrograde à cours terme jouent un rôle fondamental dans cette perception, et nous manquons d'éléments sur la mémoire rétrograde quant à son rôle sur la perception. Notre modèle de grille peut s'interpréter comme un retournement temporel des événements de par la symétrie de la grille et des points pivots ou centres de la grille considérés pour chaque événement.

Dans ce chapitre, nous présentons une solution pour estimer le suivi du tempo à travers la suite d'événements qui l'induisent. Nous présentons en (5.1.2) le modèle probabiliste issu de l'observation de nos données et la séquence par modèle de Markov caché (*HMM*) qui en résulte. En (5.1.3) nous présentons l'inférence dynamique avec un algorithme de décodage *Viterbi* appliqué à la séquence de variables cachées. Nous séparons estimation de tempo et estimation de phase que nous reconstruisons à l'aide de la fonction de *focus attentionnel* [Large and Jones, 1999] [Large and Palmer, 2002] que nous avons introduite à la section (5.2). Nous décomposons ainsi le suivi de tempo en deux temps : une inférence du tempo et une synchronisation (5.3) réalisée par un entraînement de la grille en réponse aux stimulations extérieures des événements.

5.1 Inférence et modèle probabiliste

[Sadakata et al., 2005] justifient le cadre bayésien comme une nouvelle manière de comprendre la différence entre une mesure de la perception et de la production de séquences rythmiques, et ainsi de prédire des données perceptives en fonction de données de production.

Dans le cadre d'une analyse en temps réel, une tâche commune est de prédire les observations futures étant données les observations $y_{1:t} = (y_1, y_2, \dots, y_t)$ collectées jusqu'à l'instant t présent, t est un entier et désigne un temps discret. Nous désirons calculer une bonne prédiction et connaître éventuellement un indice de confiance ou de certitude relatif à ce choix. L'idée générale est alors de calculer une probabilité de distribution sur les possibles observations futures, nous la notons $P(y_{t+h}|y_{1:t})$, où $h > 0$ désigne l'horizon, i.e. la distance dans le futur jusqu'à laquelle nous désirons faire des prédictions.

5.1.1 Modèle d'espace d'états

Dans un *modèle d'espace d'états*, nous supposons qu'il existe des états cachés sous-jacents du *monde* qui génère les observations, et que ces états cachés évoluent avec le temps. Souvent le terme *modèle d'espace d'états* implique que l'état caché est un vecteur de $\mathbb{R}^K$, mais plus généralement il signifie un modèle dynamique qui utilise des états cachés de toute nature, continue, discrète ou les deux ensemble. Le but est d'inférer l'état caché étant donné les observations jusqu'au temps présent. Si X_t représente l'état caché à l'instant t , alors nous pouvons définir notre but comme étant précisément le calcul de $P(X_t|y_{1:t})$, c'est ce que nous appelons l'état de croyance ou *belief state*. Nous pouvons mettre à jour l'état de croyance récursivement en utilisant le théorème de Bayes que nous rappellerons ci-après et détaillé en Annexe (C). Dans le cas de la prédiction, nous maintenons une probabilité de distribution sur X_t au lieu d'un simple meilleur choix, afin de refléter notre incertitude sur le véritable état du *monde*.

Un modèle d'espace d'états comporte une probabilité *a priori* $P(X_1)$ appelée *prior*, une fonction de transition d'état $P(X_t|X_{t-1})$ et une fonction d'observation $P(Y_t|X_t)$. Nous faisons l'hypothèse que le modèle est un modèle de Markov d'ordre 1, i.e. $P(X_t|X_{1:t-1}) = P(X_t|X_{t-1})$. De même nous faisons l'hypothèse que les observations sont conditionnellement Markov d'ordre 1 : $P(Y_t|Y_{1:t-1}, X_t) = P(Y_t|X_t, Y_{t-1})$. Souvent cette dernière hypothèse est simplifiée en posant $P(Y_t|Y_{t-1}, X_t) = P(Y_t|X_t)$, (Fig.5.1). Le modèle est dit invariant et homogène si les fonctions de transition et d'observation sont les mêmes à chaque instant. Nous ne nous trouvons pas dans cette hypothèse, mais localement nous pouvons l'affirmer. De plus cette hypothèse est importante pour modéliser des séquences de longueur infinie et leur convergence, ce qui

n'est pas nécessaire dans notre cas.

Représentation

Il y a de nombreuses manières de représenter les modèles d'espace d'états [Murphy, 2002], les plus courants sont les modèles de Markov cachés – *Hidden Markov Models* ou *HMM*, et les filtres de Kalman – *Kalman Filters Models* ou *KFM* [Maybeck, 1979]. Les modèles *HMM* présupposent que X_t est une variable aléatoire discrète, $X_t \in \{1, \dots, K\}$. Les modèles *KFM* présupposent que X_t est un vecteur de variables aléatoires continues, $X_t \in \mathbb{R}^K$ et que $X_{1:T}$ et $Y_{1:T}$ sont gaussiennes. Les modèles *HMM* et *KFM* peuvent être généralisés par les Réseaux de Bayes Dynamiques – *Dynamic Bayesian Network* ou *DNB* [Murphy, 2002].

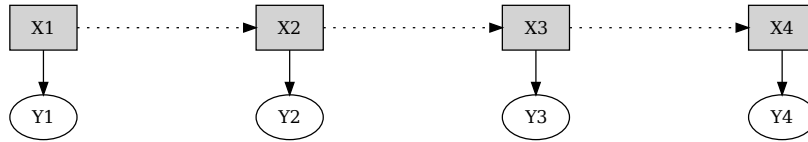


FIG. 5.1 – modèle d'état *HMM*

5.1.2 Modèle probabiliste de tempo

A l'instant t , la fenêtre des n derniers événements de la suite composée des temps d'onsets constitue une séquence d'observation temporelle $\{t_{-n+1}, t_{-n+2}, \dots, t_{-1}, t\}$ des événements. Le tempo à t représente l'état caché à l'instant t sur la fenêtre des événements considérés. La fenêtre $\{t_{-n+1}, t_{-n+2}, \dots, t_{-1}, t\}$ donne lieu après analyse par Alg.4.5 à une observation, vecteur de probabilités sur des états discrets de pas de grilles optimales susceptibles de quantifier la séquence temporelle d'entrée. Nous considérons une discrétisation du tempo sur une échelle logarithmique de base 2 pour exprimer de manière équiprobable une accélération ou une diminution de tempo. En effet, le passage de 60 bpm à 72 bpm ne représente pas le même facteur d'accélération que le passage de 48 bpm à 60 bpm, dans le premier cas il est de 1.2 soit une accélération de 20% et dans le second de 1.25 soit 25%, la même différence de 12 bpm est dépendante du tempo de départ. Une échelle en $\log_2$ rétablit une équidistance entre les accélérations de tempo. Cela nous amène à considérer les accélérations de tempo comme des rapports d'octave : si nous prenons pour exemple une *octave de tempo* de 70 à 140 bpm,

les discrétisation de tempo s'échelonnent sur les valeurs de 70, 74.2, 78.6, 83.2 ... à 140 *bpm*. Une division de l'octave de tempo en 24 est une précision suffisante et correspond à des accélérations relatives d'un facteur de $2^{\frac{1}{24}}$ soit un peu moins de 3% en *bpm* relatifs. Nous considérons alors ces divisions comme des boîtes dans lesquelles nous pouvons garder les solutions Δ des grilles optimales associés à leurs *indices de coïncidence* introduits par la fonction d'erreur normalisée $\sigma(\Delta)$ (3.3.28). Nous réalisons ainsi une quantification vectorielle du vecteur observation afin de réduire le nombre de calculs pour le décodage du tempo.

Grille optimale accélérée et densité de probabilité

Nous devons définir un cadre probabilistique qui nous permette de considérer l'ensemble des solutions pas de grille de l'Alg.4.5 telles qu'elles génèrent une fonction de densité de probabilité. Pour cela, nous cherchons une bonne estimation de la distribution de probabilité du tempo T au temps t , cette distribution est appelée *posterior*. L'Alg.4.5 renvoie pour chaque pas de grille Δ_k un *indice de coïncidence* $\sigma(\Delta_k)$ que nous pouvons interpréter comme proportionnel à la probabilité de Δ_k à représenter le tempo local, de par sa capacité à quantifier correctement la séquence d'entrée. Une autre vision de la distribution de probabilité discrète peut être présentée de la façon suivante : pour une fenêtre de n points donnés en entrée, la résolution du problème de grille de coïncidence optimale retourne un ensemble de valeurs $\{\Delta_k\}$ qui quantifient la séquence d'entrée avec une marge d'erreur $\sigma(\Delta_k)$ que nous pouvons assimiler à un histogramme sur Δ des erreurs de quantification en chaque point de la séquence.

Echantillonnage d'une distribution de probabilité

Ne connaissant pas la distribution de probabilité *posterior*, cela revient à effectuer un échantillonnage du *posterior* contrôlé par la méthode proposée de grille optimale. Pour notre modèle, cet échantillonnage est particulier car optimum de la distribution de probabilité du tempo pour réaliser l'inférence. Notre distribution cible étant la distribution de probabilité du tempo à un instant donné t nous aurions pu utiliser une méthode de type Monte Carlo avec Chaîne de Markov (*MCMC*) [Chakraborty, 2002a] [Chakraborty, 2002b] pour l'estimer. Néanmoins, nous avons démontré (Th.3.3.2, Th.3.3.6, Th.4.2.6) que les points de discontinuité de la fonction d'erreur sont les points correspondants aux solutions optimales de cette fonction et leur répartition n'est pas

uniforme.

***MCMC* et grille optimale accélérée**

Les méthodes de type *MCMC*, grâce à un théorème connu sur la convergence des Chaînes de Markov, se basent sur le fait qu'un échantillonnage correct de la distribution cible *posterior* peut être obtenu comme une valeur limite de la probabilité *posterior* échantillonnée suivant un choix d'échantillons générés par une Chaîne de Markov bien choisie et définie à partir de distributions marginales et conditionnelles données. Plusieurs problèmes se posent alors : comment estimer de manière optimale la Chaîne de Markov à chaque fenêtre d'analyse, et quelle est la qualité de convergence de la distribution limite atteinte à l'aide de la *MCMC* ? L' Alg.4.5 permet de contourner ces deux difficultés, est beaucoup plus efficace en temps de calcul et donne une estimation exacte des points importants de la distribution cible sous les hypothèses considérées. En un sens on peut y voir une analogie avec les méthodes relatives à l'*importance sampling* – échantillonnage de Gibbs ou algorithme de Metropolis [Tierney, 1994][Chib and Greenberg, 1995]. Bien sûr notre distribution est valable dans l'hypothèse où notre paradigme modélise correctement notre *monde* rythmique, ce qui est une approximation de la réalité qu'il ne faut pas oublier : nous utilisons un modèle génératif qui modélise une inversion du processus réel. Ainsi, dans ce cadre, l'algorithme de grille optimale accélérée est nécessairement meilleur qu'une *MCMC* pour l'échantillonnage de la distribution *posterior*.

Vision bayésienne de l'inférence de tempo

Sadakata, Desain, and Honing [2005] partent de la remarque générale suivante : le rapport d'une mesure – choisie et adéquate – de la perception et de la production de séquences rythmiques simples n'est pas nécessairement cohérent. Sadakata et al. montrent qu'une approche bayésienne propose une nouvelle manière d'appréhender cette différence. La formalisation de la compétition perceptive entre différentes représentations mentales est possible dans ce cadre qui présuppose alors l'existence de probabilités *a priori* non uniformes concernant les catégories rythmiques : celles-ci étant liées au tempo, nous pouvons étendre cette approche à la question du tempo. Il devient alors possible de relier les informations de perception et de production afin de prédire en conséquence des données perceptives à partir de données de production.

Cette formalisation bayésienne peut être interprétée comme une adaptation optimale de notre système perceptif à l'environnement dans lequel la production d'événements rythmiques a lieu. L'idée est ici de modéliser une forme d'écoute rythmique et d'inférer les informations à partir de données qui proviennent d'une séquence de points temporels émis à travers un canal mono-dimensionnel.

Notre problème met en oeuvre une séquence de la variable d'état cachée X qui représente le tempo. A l'instant t , le tempo est X_t et l'Alg.4.5 sur la séquence des n derniers événements $\{t_{-n+1}, t_{-n+2}, \dots, t_{-1}, t\}$ retourne un vecteur observation Y_t .

5.1.3 HMM et décodage Viterbi

Le suivi de l'estimation de tempo peut être ramené à un modèle d'état pour lequel nous utilisons un décodage Viterbi [Forney, 1973] appliqué à un modèle de Markov Caché (*HMM*). L'observation est une fonction probabiliste de l'état caché, on appelle cette probabilité le *likelihood* $P(Y|X)$: elle exprime la probabilité d'observation ou de génération des données Y par l'état X . Un modèle *HMM* est défini complètement par [Rabiner, 1989] :

- N , le nombre d'états du modèle : $S = \{S_1, S_2, \dots, S_N\}$. L'état au temps t est q_t
- M , le nombre de symboles d'observation distincts par état : $V = \{v_1, v_2, \dots, v_m\}$
- $A = \{a_{ij}\}$, la distribution de probabilité de transition d'état, avec :

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad 1 \leq i, j \leq N \quad (5.1.1)$$

$$\text{et} \quad \sum_{j=1}^n a_{ij} = 1 \quad (5.1.2)$$

Si tous les $a_{ij} > 0$ alors le modèle est ergodique.

- $B = \{b_j(k)\}$, la distribution de probabilité d'observation du symbole B dans l'état j , avec :

$$b_j(k) = P(v_k | q_t = S_j) \quad \begin{array}{l} 1 \leq j \leq N \\ 1 \leq k \leq M \end{array} \quad (5.1.3)$$

– $\pi = \{\pi_i\}$, la distribution de l'état initial, où

$$\pi_i = P(q_1 = S_i) \quad 1 \leq i \leq N \quad (5.1.4)$$

Le modèle *HMM* avec ses paramètres N, M, A, B, π se résume dans la notation compacte

$$\lambda = (A, B, \pi) \quad (5.1.5)$$

Estimation dynamique des états de tempo

Considérant le modèle *HMM* ci-dessus, les états cachés de tempo sont les boîtes à valeurs discrètes de tempo sur l'échelle $\log_2$ introduites en (5.1.2). Les symboles des observations sont également ces mêmes boîtes de tempo discrétisé. A chaque instant t , nous observons le vecteur de probabilité *likelihood* :

$$P(Y_t|X_t) = [P(y_{1,t}|X = S_1), P(y_{2,t}|X = S_2), \dots, P(y_{n,t}|X = S_n)] \quad (5.1.6)$$

Le décodage Viterbi permet avec le théorème de Bayes qu'il intègre d'estimer la probabilité *a posteriori* :

$$Posterior_{(t+1)} = Likelihood_{(t+1)} \cdot Prior_{(t)} \quad (5.1.7)$$

$$= Likelihood_{(t+1)} \cdot \alpha_t \cdot Transition_{(t \rightarrow t+1)} \quad (5.1.8)$$

$$P(X_{t+1}|Y_{1:t+1}) = P(Y_{t+1}|X_{t+1}) \cdot \alpha_t \cdot Transition_{(t \rightarrow t+1)} \quad (5.1.9)$$

avec α défini récursivement par [Rabiner, 1989] :

$$\alpha_{t+1}(j) = \sum_{i=1}^N [\alpha_t(i) a_{ij}] y_{j,t+1} \quad (5.1.10)$$

Ce faisant, dans (5.1.10), nous remplaçons le terme $b_{j,t+1}(k)$ de (5.1.3) par l'observation locale $y_{j,t+1}$, ce qui rend la distribution de probabilité d'observation B non stationnaire. La matrice de probabilité de transition est choisie de manière à autoriser les sauts de tempo, elle est définie ainsi

$$a_{ij} = \mathcal{N}(j, \mu_{[=i]}, \varepsilon_{[\infty|i-j]]) \quad (5.1.11)$$

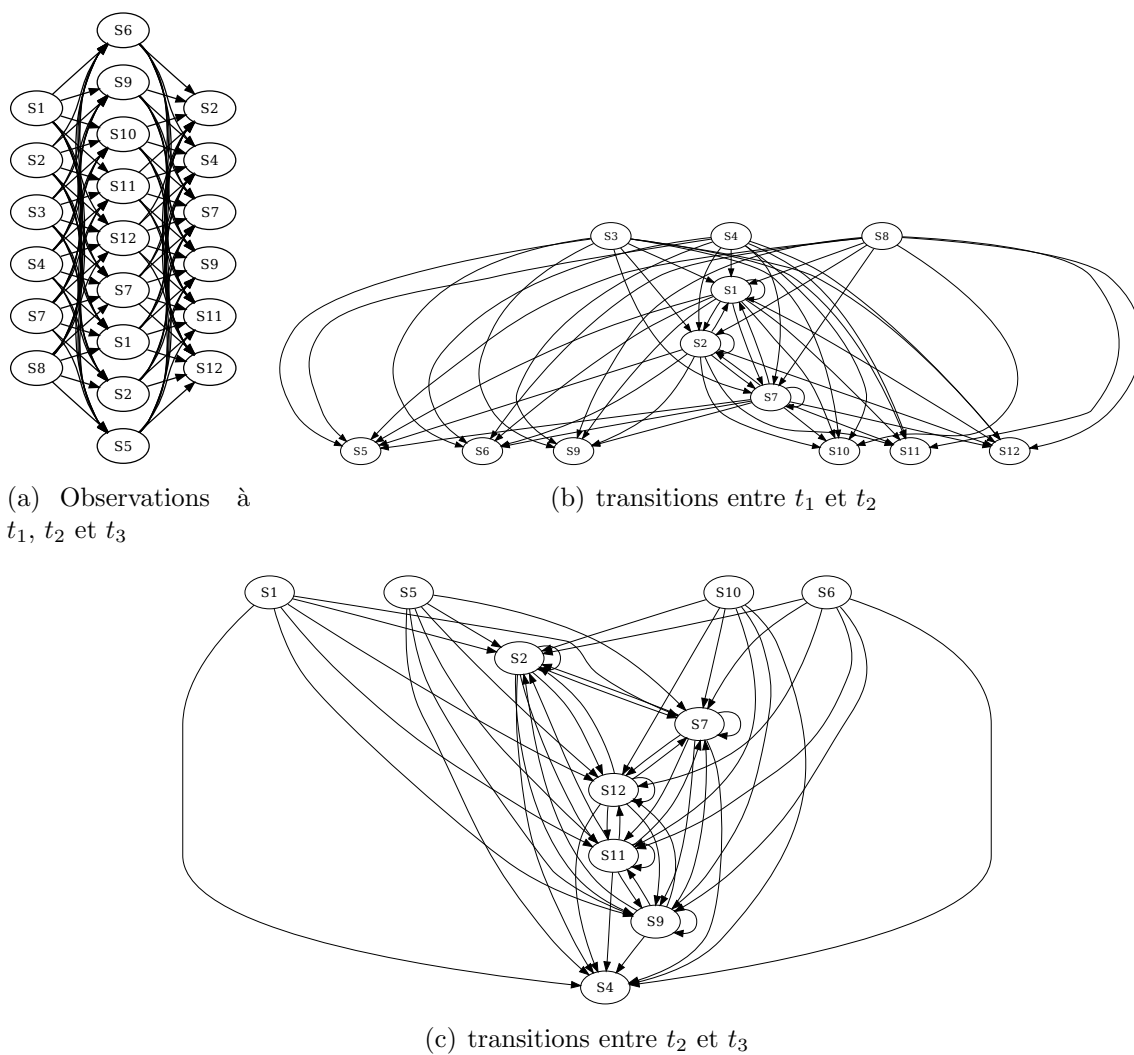


FIG. 5.2 – exemple d'observations et transitions d'états associées

Avec $\mathcal{N}(\cdot, \mu, \varepsilon)$ fonction de probabilité gaussienne de moyenne μ et de variance ε . En jouant sur ε proportionnel à $|i - j|$, on peut devenir très sélectif sur l'évolution du tempo ou bien au contraire se rapprocher d'une distribution uniforme qui ne pénalise pas les sauts de tempo. A l'instant $t + 1$, la distribution *posterior* (5.1.9) de la variable cachée X reflète notre entière connaissance sur le tempo, i.e. les valeurs probables et les incertitudes associées. Nous pouvons obtenir l'état le plus probable x^* de X connaissant Y :

$$x^* = \underset{x}{\operatorname{argmax}} P(x|y) \quad (5.1.12)$$

Conclusion

A l'aide de l'algorithme de Viterbi pour décoder les observations Y , nous obtenons en temps réel à l'instant t et à chaque itération l'estimation de tempo la plus probable. Ne pouvant utiliser les informations futures en raison de la causalité du système, nous calculons uniquement la partie α directe - ou *forward* - de l'algorithme de *Viterbi*. La figure Fig.5.3 montre un exemple de chemin de tempo inféré par la résolution dynamique effectuée avec l'algorithme de *Viterbi* sur un fichier midi issu d'une analyse du signal audio.

5.2 Synchronisation : un modèle d'oscillateur non linéaire

Large and Jones [1999] propose certainement le modèle de suivi de tempo le plus intéressant du point de vue de la théorie en rapport avec le modèle cognitif développé. Une théorie de dynamique attentionnelle est proposée, l'approche décrit l'écoute active liée à un comportement d'oscillations internes - les rythmes couplés - capables de se synchroniser sur des événements extérieurs et de focaliser une énergie attentionnelle en certains moments temporels attendus. Il est ainsi possible de suivre dans une certaine mesure des événements issus de rythmes complexes. Comme il le précise, Il y a en un élément fascinant de flexibilité dans la structure temporelle d'événements naturels, et notamment dans la temporalité musicale, où des observateurs peuvent appréhender des structures rythmiques stables quand bien même les périodicités qui composent ces structures fluctuent grandement. Large and Jones [1999] part de l'obser-

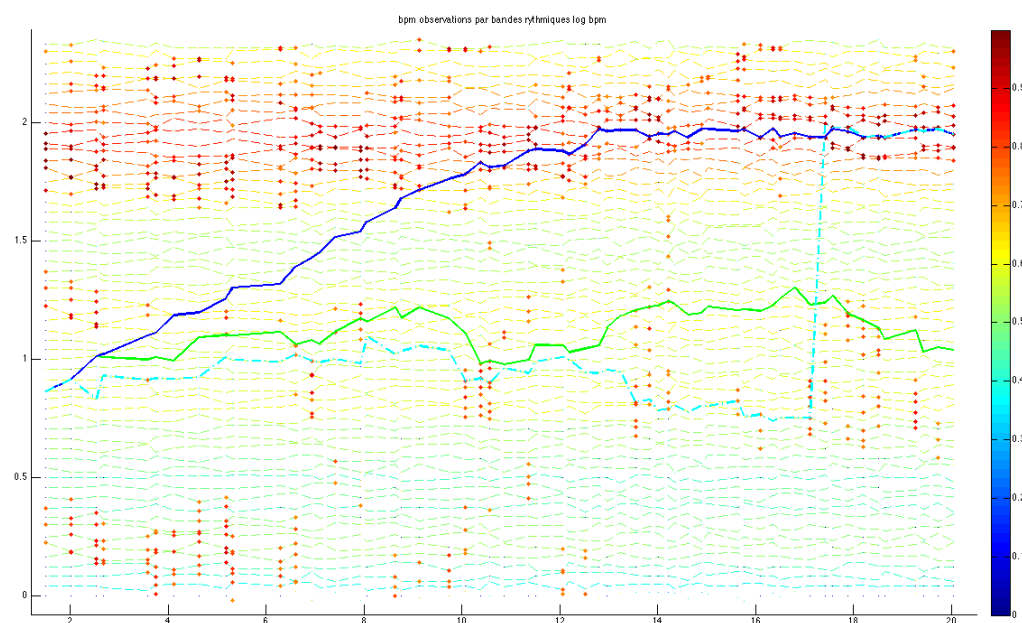


FIG. 5.3 – inférences du tempo

variation des intervalles de temps entre les événements – inter-onset – et montre que dans le domaine auditif, la sélectivité attentionnelle opère dans la dimension temporelle. Les deux questions principales auxquelles il tente de répondre sont les suivantes : comment des structures d'événements guident l'attente, en renforçant sélectivement l'attention en des instants particuliers ? Et comment l'attente s'adapte-elle quand la structure événementielle commence à se modifier, dans le sens où des informations importantes apparaissent à des instants inattendus ? *théorie des structures temporelles* , *theorie statistique et beat-based theories* Finalement, la question qui articule la prédiction du tempo est celle-ci : quand est-ce que la différence entre deux intervalles est-elle assez importante pour induire un sens de mètre ou de rythme, et quand est-ce qu'elle est assez faible pour être considérée comme une simple perturbation d'une séquence isochrone ? Des éléments des deux approches, statistiques et sur la pulsation sont nécessaires, et plus qu'une simple mémoire de structure temporelle, c'est l'écoute, l'attente active qui est sollicitée.

Oscillateur non linéaire

La formalisation mathématique proposée par [Large and Jones \[1999\]](#) a pour base un oscillateur auto-entretenu qui modélise le processus de dynamique attentionnelle. Il génère premièrement une activité périodique liée à l'attente, considérée comme une anticipation implicite. Couplé à une source rythmique extérieure, l'oscillateur peut se synchroniser par entraînement à cette source temporelle.

C'est la phase $\Phi(t)$ qui permet de décrire *quand* un événement doit arriver, elle correspond à la position de l'oscillation autour du cycle stable de l'oscillateur au temps t . Du point de vue de la terminologie des systèmes dynamiques, la phase est un des trois paramètres du modèle et

$$\Phi(t) = \frac{t - t_x}{p} \quad , \quad t_x - \frac{p}{2} \leq t < t_x + \frac{p}{2} \quad (5.2.1)$$

p étant la période de l'oscillateur et t_x le temps auquel un événement est supposé arriver. Remarquons qu'une attente correspond toujours à une phase nulle. $\Phi(t)$, ou simplement Φ capture la disparité entre un onset et sa prédiction. Une phase négative correspond à un événement en avance par rapport à celui attendu, et une phase positive à un événement en retard. Il est alors possible d'exprimer (5.2.1) sous une forme récursive de l'onset n vers l'onset $(n + 1)$:

$$\Phi_{n+1} = \Phi_n + \frac{t_{n+1} - t_n}{p_n} \quad (mod_{-0.5}^{+0.5}) \quad (5.2.2)$$

(5.2.2) modélise la relation de deux rythmes qui sont découplés, mais ne modélise pas la force exercée par un rythme extérieur sur un rythme attentionnel. Cette force peut être modélisée en ajoutant un terme de couplage $F(\Phi_n)$:

$$\Phi_{n+1} = \Phi_n + \frac{t_{n+1} - t_n}{p_n} - \eta_\Phi F(\Phi_n) \quad (mod_{-0.5}^{+0.5}) \quad (5.2.3)$$

où $F(\Phi_n) = \frac{1}{2\pi} \sin(2\pi\Phi)$, η_Φ représente la force de couplage, (5.2.3) est dénommée carte circulaire d'attraction de phase, elle représente une forme de synchronisation souple basée sur la coordination des phases du rythme attentionnel et de l'oscillateur. $F(\Phi_n)$ sera par la suite modifié par une modélisation du *focus attentionnel* décrit ci-après.

Afin de maintenir une synchronicité de phase avec des événements extérieurs, la période

de l'oscillateur peut s'ajuster d'une manière similaire à l'ajustement de phase :

$$p_{n+1} = p_n + p_n \eta_p F(\Phi_n) \quad (5.2.4)$$

avec (5.2.2) la période devient une variable d'état, le taux d'adaptation de la période, η_p est analogue à la force de couplage de phase η_Φ de (5.2.3), ces deux paramètres déterminent ensemble le taux d'approche de l'attracteur.

focus attentionnel

Large and Palmer [2002] complète ce modèle avec la notion d'attention focalisée — *focus attentionnel*, une troisième variable d'état décrit la concentration de l'énergie attentionnelle pour chaque cycle de l'oscillateur. Le *focus* attentionnel accomode les effets d'attente temporelle sur la réponse observée. Dans une première recherche, Large [1994] modélise la pulsation d'énergie attentionnelle de manière suivante :

$$o(t) = 1 + \tanh \gamma (\cos 2\pi\Phi - 1) \quad (5.2.5)$$

$$\text{avec } \gamma = \frac{-0.416}{\cos 2\pi\tau - 1} \quad (5.2.6)$$

où γ est un gain, lié à τ qui s'interprète comme une mesure de réceptivité dans le domaine temporel, i.e. une mesure du *focus attentionnel* illustré avec la Fig. 5.4.

La nature de non distribution de (5.2.5) conduit Large and Palmer [2002] à chercher une meilleure pulsation pour décrire l'énergie attentionnelle. Il la modélise par une fonction de densité de probabilité périodique, la pulsation étant liée au système dynamique des équations (5.2.3) et (5.2.4) à travers la phase Φ du rythme observé. La variable d'état κ est introduite pour capturer le *focus* attentionnel, elle comporte un lieu pour $\Phi = 0$ — moment de la pulsation — ainsi qu'une zone d'extension dans le temps qui détermine la largeur du *focus*.

$$f(\Phi, \kappa) = \frac{1}{I_0(\kappa)} e^{\kappa \cos 2\pi\Phi} \quad (5.2.7)$$

$I_0(\kappa)$ est la fonction de Bessel de première espèce d'ordre 0 modifiée. (5.2.7) correspond à la fonction de densité de probabilité périodique de von Mises, elle décrit une inférence statistique sur un cercle, on peut la comprendre comme une gaussienne dans le domaine circulaire. Fig. 5.5 représente la distribution de von Mises en coordonnées cartésiennes et polaires, avec des valeurs de κ variant de 0 à 1. De même que pour les autres variables

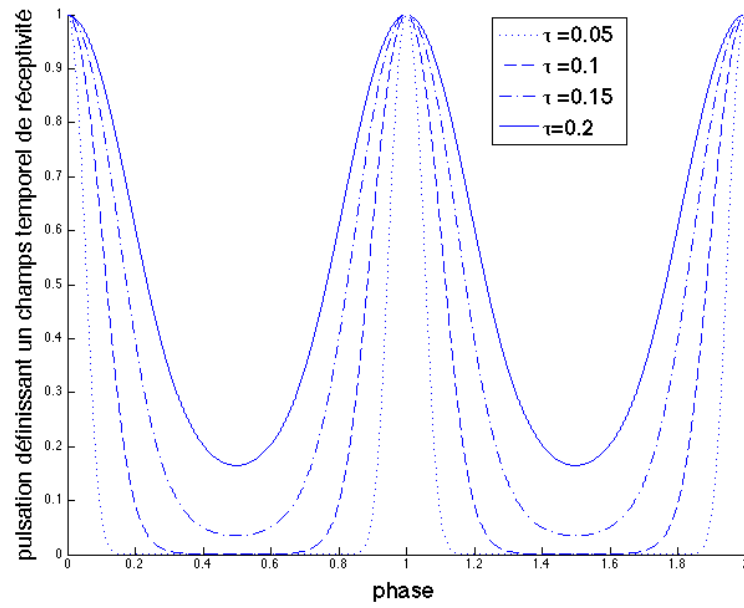


FIG. 5.4 – Pulsation modélisant un champs de receptivité temporel

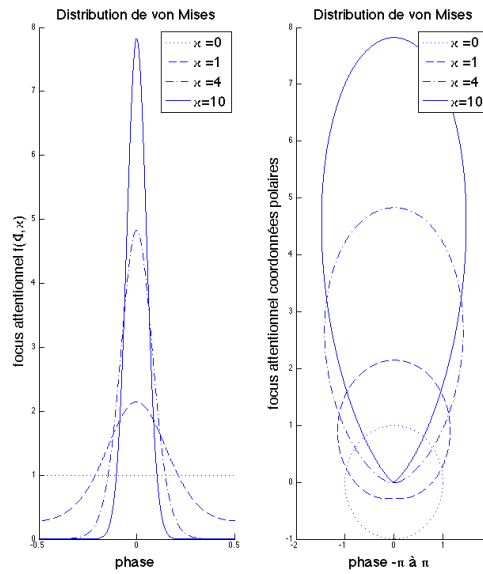


FIG. 5.5 – Distribution de von Mises, coordonnées cartésiennes et polaires

d'état, κ est adapté à chaque pas, par une méthode d'estimation du *maximum likelihood* de κ pour la distribution de von Mises. Ce maximum est donné par la solution de l'équation :

$$A(\kappa) = r \quad (5.2.8)$$

où r est le *module du vecteur moyen* obtenu dans la statistique du domaine circulaire par :

$$r = \frac{1}{n} \sum_{i=1}^n \cos 2\pi(\Phi_i - \bar{\Phi}) \quad (5.2.9)$$

r peut être utilisé comme mesure de synchronisation et varie de 0 à 1, où 0 indique qu'il n'y a pas synchronisation et 1 indique une synchronisation parfaite. r est comparable à une forme de variance dans le domaine circulaire, il représente une mesure de dispersion. A est une fonction monotone de κ telle que, quand r varie de 0 à 1, κ varie de 0 à $+\infty$. Avec (5.2.8), l'estimation du *maximum likelihood* de κ – paramètre du modèle probabiliste de von Mises, devient alors :

$$A(\hat{\kappa}) = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{1}{n} \sum_{i=1}^n \cos 2\pi\Phi_i \quad (5.2.10)$$

la partie droite de (5.2.10) étant le vecteur moyenne statistique r , celui-ci est approché de manière incrémentale par la suite $\{s_n\}$:

$$s_n = s_{n-1} - \eta_s(s_{n-1} - \cos 2\pi\phi_n) \quad (5.2.11)$$

car on peut montrer sans faire d'approximation que

$$s_n - s_{n-1} = -\frac{1}{n}(s_{n-1} - \cos 2\pi\phi_n) \text{ avec } s_0 = 0 \quad (5.2.12)$$

η_s apparaît donc dans (5.2.11) comme un facteur de lissage sur les n derniers événements. L'inversion de (5.2.10) donne alors κ :

$$\kappa_{n+1} = A^{-1}[H(b, s_{n+1})] \quad (5.2.13)$$

A^{-1} est résolu par approximation numérique ou par lecture de table, car la fonction A n'est pas inversible analytiquement, b est une limite supérieure telle que la valeur maximale de κ soit $A^{-1}(b)$, la limite du *focus attentionnel* est fixée à $b = 0.95$. Le taux d'adaptation $0 \leq \eta_s \leq 1$ détermine une constante de temps d'adaptation, avec de faibles valeurs approxinant r sur un temps plus long. Le modèle de *focus attentionnel* développé est alors intégré dans le terme de couplage $F(\Phi_n)$ introduit en (5.2.3) qui devient $F(\Phi_n, \kappa)$. Celui-ci est maintenant la dérivée d'une version unitaire normalisée

de la pulsation attentionnelle et nous avons $F(\Phi_n, 0) = F(\Phi_n)$:

$$F(\Phi_n, \kappa_n) = \frac{1}{2\pi \exp \kappa} [\exp \kappa \cos 2\pi \Phi] \sin 2\pi \Phi \quad (5.2.14)$$

La **Fig.5.6** illustre la relation du *focus attentionnel* avec la correction de phase associée pour des valeurs de κ variant de 0 à 10. Si la phase est négative, la correction se fait dans le sens opposé.

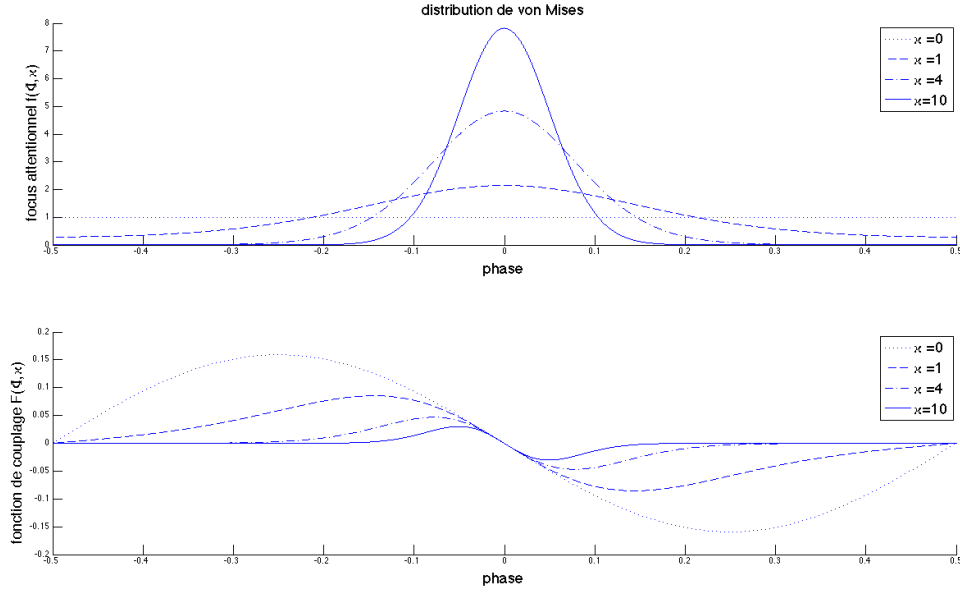


FIG. 5.6 – Focus attentionnel et fonction de couplage attentionnel

l'équation (5.2.3) est alors :

$$\Phi_{n+1} = \Phi_n + \frac{t_{n+1} - t_n}{p} - \eta_\Phi F(\Phi_n, \kappa_n) \quad (\text{mod}_{-0.5}^{+0.5}) \quad (5.2.15)$$

finalement, les équations (5.2.3) et (5.2.4) deviennent :

$$\Phi_{n+1} = \Phi_n + \frac{t_{n+1} - t_n}{p_n} \quad (5.2.16)$$

$$\hat{\Phi}_{n+1} = \Phi_{n+1} - \eta_\Phi \frac{1}{2\pi e^\kappa} e^{\kappa \cos 2\pi \Phi_{n+1}} \sin 2\pi \Phi_{n+1} \quad (\text{mod}_{-0.5}^{+0.5}) \quad (5.2.17)$$

$$p_{n+1} = p_n \left(1 + \eta_p \frac{1}{2\pi e^\kappa} e^{\kappa \cos 2\pi \Phi_{n+1}} \sin 2\pi \Phi_{n+1} \right) \quad (5.2.18)$$

Ainsi, le comportement de l'oscillateur non linéaire est déterminé par ses 3 variables

d'état Φ , p et κ , et les 3 paramètres respectifs associés, la force de coupage de phase η_Φ , le taux d'adaptation de la période η_p et le taux d'adaptation du *focus attentionnel* η_κ , respectant en général $0 < \eta_\kappa < \eta_p < \eta_\Phi \leq 1$. L'algorithme [Alg.5.1](#) résume le processus d'adaptation incrémental :

1. Dans un premier temps, [\(5.2.19\)](#) ou [\(5.2.20\)](#) évalue la phase Φ_n de l'onset t_n qui arrive par rapport au mode courant de l'oscillateur – i.e. le dernier instant auquel sa phase a été nulle.
2. On adapte le *focus attentionnel* κ_n de la prochaine pulsation avec Φ_n dans [\(5.2.22\)](#).
3. On prédit la prochaine pulsation à l'aide de l'adaptation de phase [\(5.2.23\)](#), elle a pour probabilité de champs de réceptivité la distribution de von Mises paramétrée par κ_n . On prédit aussi le nouveau tempo avec [\(5.2.24\)](#).

Le cycle de *mesure* (1), *correction* (2) et *prédiction* (3) s'apparente à un algorithme de filtrage de Kalman étendu, [Maybeck \[1979\]](#).

[Cont \[2008a\]](#) utilise une version comparable de [Alg.5.1](#) pour intégrer le tempo dans *Antescofo*, suivi de partition anticipatif : un agent temporel de décodage de tempo agit en collaboration avec un agent de décodage audio afin d'assurer un suivi robuste de partition qui a été utilisé dans plusieurs situations et productions musicales [Harvey et al. \[2008\]](#), [Stroppa and Cont \[2008\]](#). La différence principale porte sur la phase Φ_n qui est calculée par rapport à une phase pré-générée issue de la partition symbolique à suivre.

5.3 Synchronisation et prédiction de la pulsation

Nous posons l'hypothèse de manière indépendante la synchronisation de la pulsation anticipée. Le suivi de tempo dynamique implique deux tâches distinctes : le suivi de l'estimation du tempo et le suivi, ou la reconstruction, de la phase du tempo. Il est vrai que les deux problématiques sont liées de manière structurelle et fonctionnelle. [\[Large, 1995\]](#) le démontre avec un modèle d'oscillateurs non-linéaires couplés : la phase et le tempo sont couplés de manière optimale par son modèle, mais le modèle est contraint à une certaine continuité due au terme de couplage [\(5.2.3\)](#) qui apparaît avec les paramètres de force de couplage de phase η_Φ et de taux d'adaptation de la période η_p . Le problème se pose alors chez [\[Large, 1995\]](#) de pouvoir faire intervenir une

Algorithm 5.1 modèle d'oscillateur non linéaire avec *focus attentionnel*

données :

t_{n-1} instant de l'événement $n - 1$,
 Φ_{n-1} phase à t_{n-1} ,
 p_{n-1} tempo en *spp* (seconde par pulsation) à t_{n-1} ,
 s_{n-1} vecteur moyenne statistique à t_{n-1} ,
 t_n instant de l'événement n

1: *mesure (1) :*

$$\Phi_n \leftarrow \Phi_{n-1} + \frac{t_n - t_{n-1}}{p_{n-1}} \quad \text{phase courante} \quad (5.2.19)$$

$$\text{ou bien } \Phi_n \leftarrow \frac{t_n - t_x}{p_{n-1}} \quad \text{avec } t_x \text{ dernier instant où } \Phi = 0 \quad (5.2.20)$$

2: *correction (2) :*

$$s_n \leftarrow s_{n-1} - \eta_s (s_{n-1} - \cos 2\pi\phi_n) \quad (5.2.21)$$

$$\kappa_n \leftarrow A^{-1} [H(b, s_n)] \quad \text{indice de } \textit{focus attentionnel} \quad (5.2.22)$$

3: *prédiction (3) :*

$$\Phi_n \leftarrow \Phi_{n+1} \leftarrow \Phi_n - \eta_\Phi \frac{1}{2\pi e^{\kappa_n}} e^{\kappa_n \cos 2\pi\Phi_n} \sin 2\pi\Phi_n \quad (\text{mod}_{-0.5}^{+0.5}) \quad (5.2.23)$$

$$p_n \leftarrow p_n \left(1 + \eta_p \frac{1}{2\pi e^{\kappa_n}} e^{\kappa_n \cos 2\pi\Phi_n} \sin 2\pi\Phi_n \right) \quad (5.2.24)$$

4: **return** Φ_n, p_n

discontinuité, un saut de tempo qui réponde à une réalité musicale. Pour cela, [Large, 2001] utilise un réseau d'oscillateurs en parallèle, chacun comprenant ses paramètres adaptatifs internes. Néanmoins, c'est plus une modélisation de la structure métrique qui ressort du réseau d'oscillateurs, malgré une compétition entre oscillateurs activés à partir d'un certain seuil relatif à l'énergie de leurs paramètres. En théorie un grand nombre d'oscillateurs couplés peut être utilisé, mais se pose alors le problème pratique de leurs gestions mutuelles.

Large adapte le tempo et la phase du tempo de manière couplée. Nous proposons ci-après un découplage de l'adaptation du tempo – réalisé par le suivi dynamique de la grille optimale – et de la phase de la pulsation – qui utilise alors une moitié du modèle de [Large and Palmer, 2002]. Une caractéristique très intéressante du modèle d'oscillateur non linéaire est due à l'introduction de la fonction de *focus attentionnel* κ qui engendre la non-linéarité adaptative de l'oscillateur. Le *focus attentionnel* permet de réguler de manière contrôlée l'adaptation de phase en fonction de la valeur de la période de l'oscillateur. La figure Fig.5.7 montre l'adaptation ou la correction de phase avec différentes valeurs de la force de couplage η . Aux cotés de l'estimation de suivie du tempo réalisée à l'aide de la *grille optimale*, nous représentons la phase du tempo à l'aide d'un oscillateur non-linéaire. La phase est ainsi traitée indépendamment du suivi : nous la reconstruisons aussi de façon indépendante en la déroulant, avec la particularité d'effectuer simultanément une synchronisation qui tient compte de la mise à jour du tempo lors de ce déroulement de phase. Dans le rôle du rythme attentionnel sujet d'observation de l'oscillateur non-linéaire nous considérons les n événements de la fenêtre glissante d'observation. Le dernier événement arrivé déclenche alors une mise à jour du tempo par la propagation dynamique de la *grille optimale* puis une mise à jour de la phase par rapport à la pulsation interne de l'oscillateur dans sa nouvelle configuration de tempo. La pulsation interne de l'oscillateur constitue alors la prédiction de la pulsation qui synchronise les événements observés et la grille choisie. En d'autres mots, nous utilisons l'inertie de la pulsation attentionnelle pour synchroniser la phase de la pulsation émise avec le rythme observé à travers la fonction de *grille optimale* qui renseigne l'estimation de tempo.

Remarquons que l'intérêt de dérouler la phase à l'aide de la fonction *kappa* pour la synchronisation assure de ne pas sauter de pulsation : si la phase est en retard, elle restera toujours en retard, de même si elle est en avance. Ce problème paraît relativement trivial mais se révèle être très difficile à respecter : il y a toujours une pulsation qui peut être oubliée avec un système qui recalerait la pulsation uniquement

grâce à des règles de comparaison sur la prédiction de phase.

La Fig.5.8 montre une mesure de la synchronisation de la pulsation avec les onsets entrants. La fonction κ devient un indice de bonne synchronisation : plus κ est grand et meilleure est la synchronisation. Nous pouvons voir sur cette figure la phase de la pulsation avec certains point de décrochage dû à un saut de tempo ou une adaptation plus difficile, néanmoins, la pulsation reste relativement fluide à l'écoute. La phase de la pulsation attentionnelle est multipliée par la fonction κ .

Analogie avec un filtre de Kalman

L'algorithme de [Large and Jones, 1999] fait intervenir une fonction κ que l'on peut assimiler à une variance. Le cycle se déroule suivant trois étapes de *mesure* (1), *correction* (2) et *prédiction* (3). Le modèle de Edward Large met en oeuvre une probabilité gaussienne circulaire. Dans un contexte de programmation dynamique on peut comparer cette approche avec un filtre de Kalman étendu [Hainsworth and Macleod, 2004] [Maybeck, 1979], qui prend une mesure de la quantité à renseigner, met à jour les paramètres de probabilité et la propage pour une nouvelle estimation dans le futur. Dans un cadre non-causal, le filtrage de Kalman permet également de faire du *smoothing* en revisitant l'estimation des prédictions dans le sens non-causal [Maybeck, 1979].

5.4 Implémentation temps réel

Dans l'environnement *MaxMSP* ¹ et *Pd* ² nous proposons l'objet *Antepulse*. Cet objet est écrit avec *flex* ³ pour la compatibilité avec les deux environnements. *Antepulse* est basé sur l'algorithme de *grille optimale* accélérée et le décodage proposé ci-dessus (Fig.5.9). *Antepulse* répond à un flux de données qui provient d'une entrée midi ou audio convertie à l'aide d'une segmentation *on-line* – pour créer des attaques de notes ou temps *onsets*. L'utilisation principale d'*Antepulse* est de générer un signal de synchronisation en phase avec les événements rythmiques et le tempo perçu. La génération d'interaction rythmique est alors possible et l'ordinateur génère par son écoute un flux

¹<http://www.cycling74.com>

²<http://www.PureData.org>

³<http://www.pd-felxt.com>

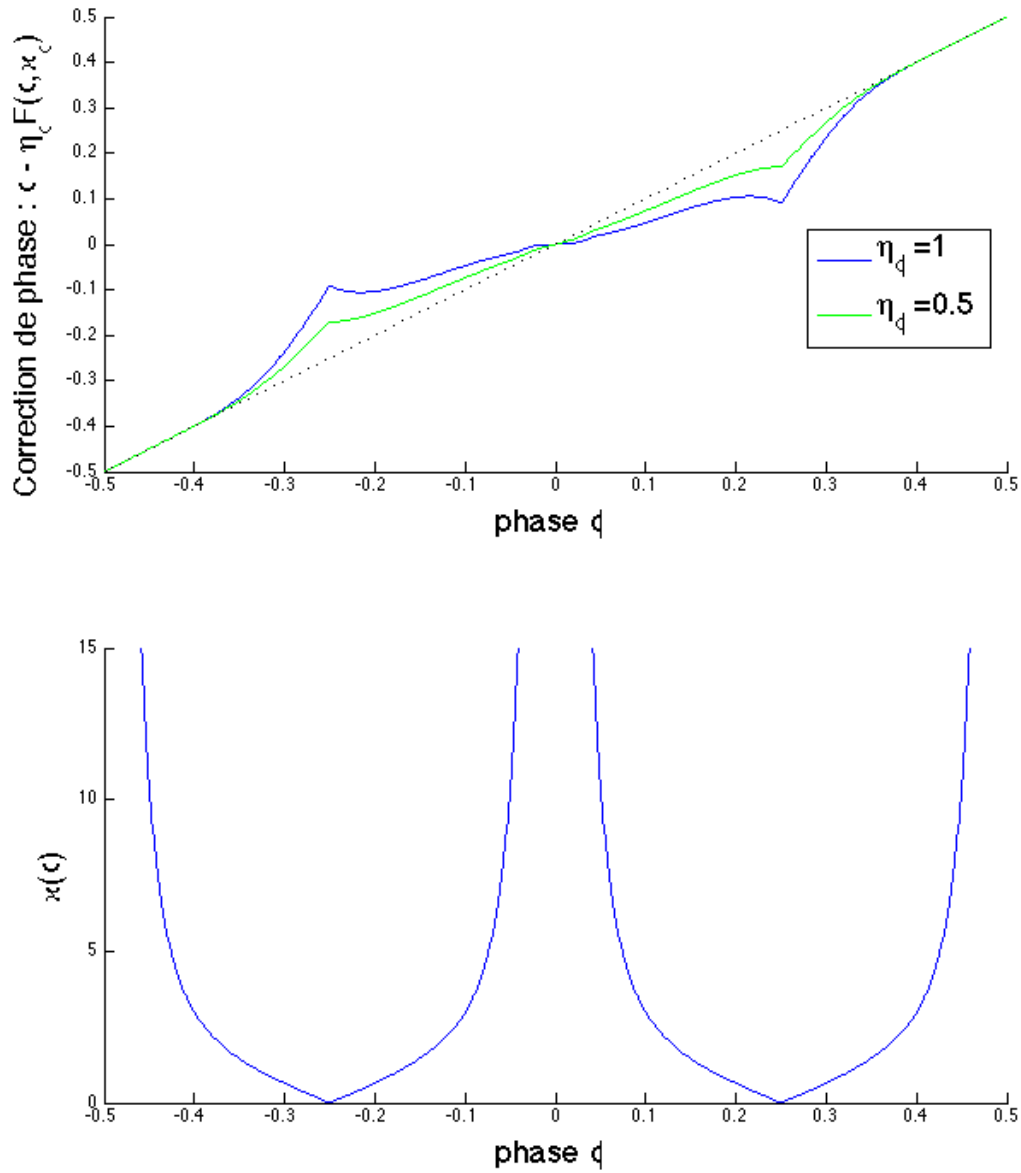
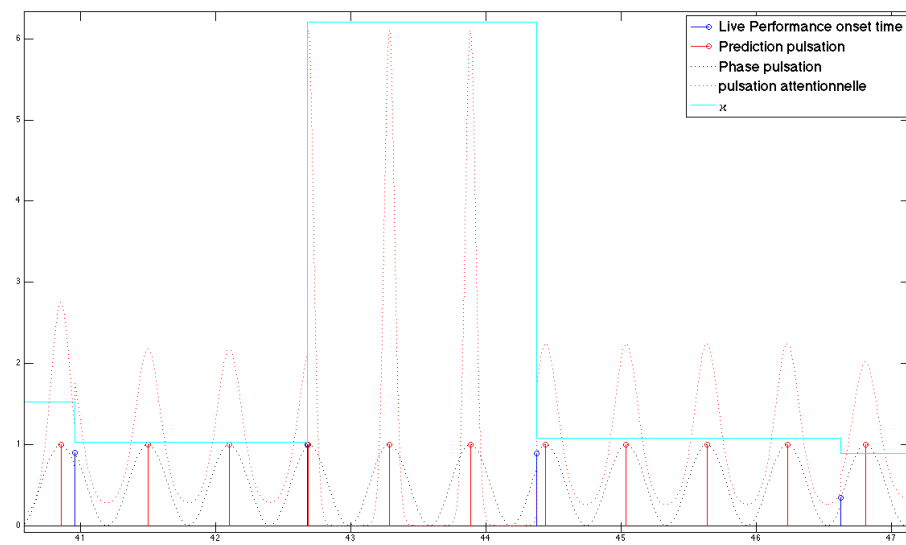
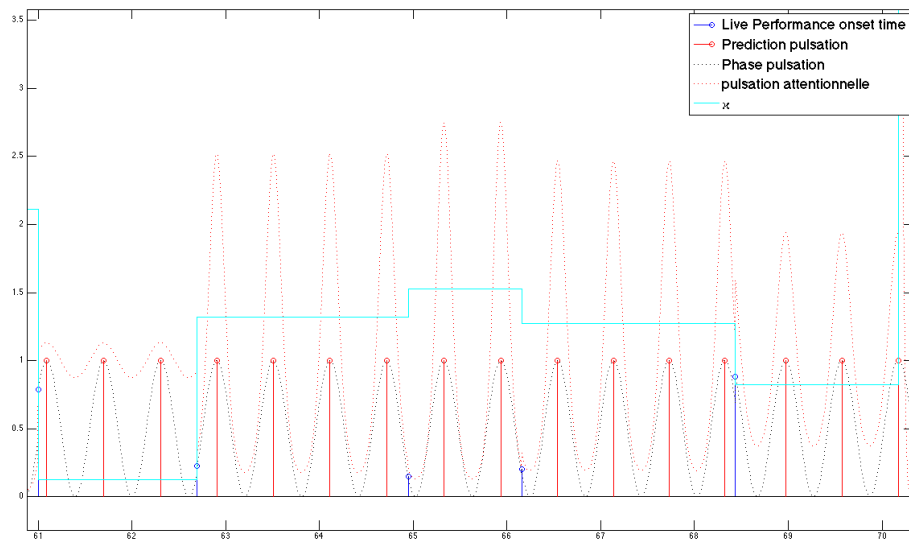


FIG. 5.7 – (ci-dessus) Adaptation de phase de la pulsation attentionnelle suivant la force de couplage η , (ci-dessous) le terme de focus attentionnel $\kappa(\Phi)$ associé.

rythmique modulé directement par ce qu'il entend. Ainsi, du point de vue de l'ordinateur qui écoute sans aucun apprentissage, avec les éléments algorithmiques donnés et les méthodes d'inférence présentées, la musique est toujours improvisée. Aucun élément stylistique ou de partitions ne sont ici disponibles afin d'orienter l'estimation



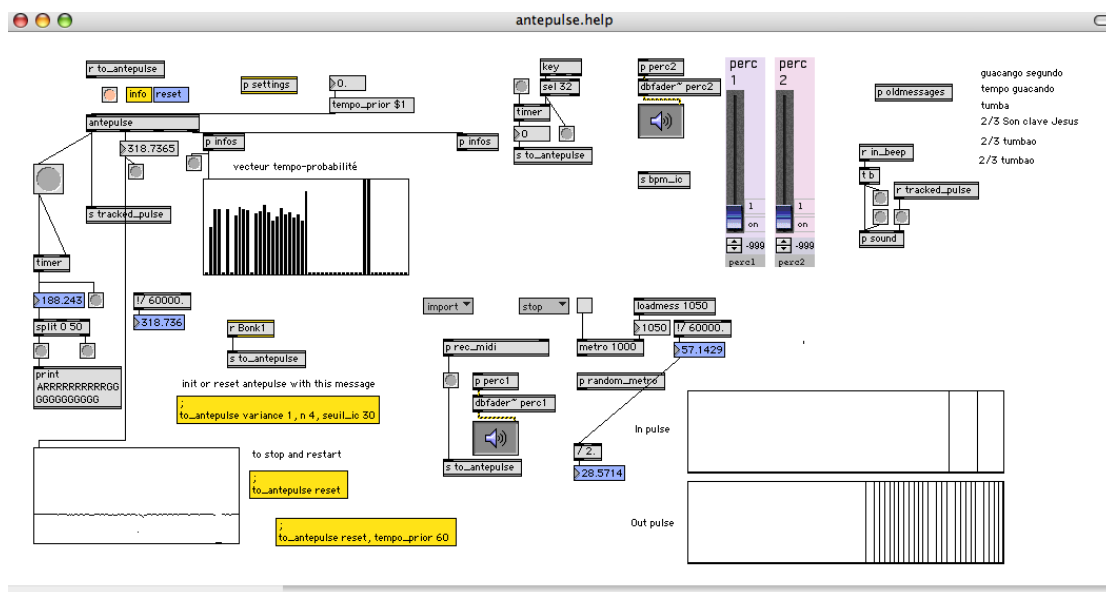
(a) Détail du déroulement de la phase du tempo avec focus attentionnel élevé



(b) détail du déroulement de la phase du tempo avec focus attentionnel plus faible

FIG. 5.8 – Adaptation de phase de la pulsation prédite. En rouge la pulsation prédite, en bleue les *onsets* en entrée

et la prédiction du tempo.

FIG. 5.9 – antepulse dans l’environnement temps réel *MaxMsp*

5.5 Expérimentations et discussion

L'algorithme de *grille optimale* est implémenté dans l'environnement temps réel *Pd* ou *MaxMSP*. Il permet de réaliser et de tester le *suivi de tempo* sur des informations musicales provenant soit de fichiers midi ou directement de capteurs reliés à un ou des instruments, par exemple des capteurs sur des éléments physiques de batterie.

Expérimentations élémentaires

Elles ont pour but de tester le modèle. Parmi les expérimentations relativement simples, nous prenons premièrement une pulsation stable que nous pouvons perturber par ajout d'un bruit aléatoire. Nous prenons également une accélération de la pulsation que nous brouillons aléatoirement. Une évaluation simple peut être de comparer la pulsation prédite à la pulsation utilisée. Néanmoins, il y a nécessité de créer une boucle de validation passant par des musiciens afin de valider la perception du tempo comme élément de comparaison de la pulsation induite.

Expérimentations sur des corpus musicaux

La validation de nos résultats devra passer par un choix de différents corpus musicaux – comme la musique contemporaine, le jazz, ou des musiques ethniques. Cette validation ne peut faire abstraction de la culture qui constitue le contexte de l’expérimentation. Néanmoins, la procédure algorithmique utilisée est celle décrite précédemment, qui ne présuppose pas de connaissances *a priori*.

La mise en place de cette validation n’a pas été effectuée dans le cadre de cette thèse et reste une extension de la recherche. Elle constitue un travail important dans un cadre qui se rapproche de la psychologie cognitive et de l’ethnomusicologie. Elle nécessite également un volume de données conséquent. En effet, il sera alors important et pertinent de vérifier en quoi l’absence de prise en compte d’indices culturels permettra ou non une déduction valide du tempo. Nous avons expérimenté de manière assez concise cette réflexion à l’aide d’enregistrements de percussions cubaines réalisées à Cuba lors d’un précédent séjour.

Nous pouvons déjà nous interroger sur la perception de la pulsation implicite et culturelle par un ordinateur qui n’a *a priori* pas d’autres éléments culturels que les hypothèses proposées, à savoir la régularité supposée induite par le tempo. Nous retrouvons alors la problématique de la transcription, liée à un facteur d’échelle temporelle, à savoir quelle notation serait utilisée par un musicien qui entendrait pour la première fois cette musique ? Cette question reste bien sûr ouverte, car la culture musicale apparaît elle aussi comme une part importante, voire fondamentale, de la perception de la pulsation et *a fortiori* du rythme. En ce sens, notre modèle cherche les limites d’une perception quasi acculturée, et envisagera par la suite, si nécessaire et suivant le contexte, des compléments d’informations culturelles pour lesquels il faudra trouver un mode de représentation adéquat. Nous présentons dans le chapitre suivant un ensemble de propositions concernant les travaux futures de notre thèse.

Chapitre 6

Conclusions et extensions futures

6.1 Conclusions des recherches en cours

Dans cette thèse, nous avons proposé une méthode *temps réel* de *suivi de tempo* d'un flux de données musicales. Nous prenons en compte seulement les temps d'apparition des événements sonores. Nous avons présenté dans un premier temps une méthode d'estimation du tempo sur un modèle de grille de pulsation optimale quasi stable puis accélérée dans un second temps. Le suivi de l'évolution de tempo, ou inférence du tempo, est réalisé par une méthode de programmation dynamique. Nous prédisons la pulsation localement et reconstruisons sa phase à l'aide d'un modèle d'oscillateur non linéaire et d'une fonction de *focus attentionnel* issues d'une modélisation cognitive. Le déroulement de phase ainsi réalisé maintient l'évolution musicale de la pulsation anticipée afin qu'elle accompagne le mieux possible le tempo suggéré par le jeu du musicien. L'idée principale liée à la perception du tempo est de toujours coller le mieux possible au temps attendu de la pulsation, d'où la notion de *meilleure coïncidence* qui sous-tend la recherche de la grille, liée à la notion perceptive de *focus attentionnel*. L'adhérence plus ou moins forte des flux temporels entre eux, que l'on trouve dans le jeu interactif des musiciens, caractérise l'élasticité du temps qui le rend spécialement musical. C'est aussi le glissement du temps symbolique et du temps de la performance, l'un par rapport à l'autre, que nous avons introduit dans le cadre de l'accélération

de tempo. Il s'agit là d'un essai de formalisation des flux temporels musicaux afin de comprendre et d'expliquer le rapport existant dans une interprétation qui met en scène plusieurs musiciens.

Nous avons centré notre recherche sur la dimension temporelle, en ne prenant pas en compte les autres paramètres sonores bien que leurs contrastes et leurs oppositions dans leurs domaines respectifs soient bien là des qualités qui participent à l'élaboration du tempo dans toute sa musicalité. Nous avons réduit le suivi de tempo à sa structure la plus dépouillée et la plus essentielle qui caractérise la nature du tempo : le temps, même si elle ne permet sans doute pas dans tous les cas d'assurer à elle seule un décodage du tempo à tout moment. Cette approche, même si elle est réductrice dans la recherche du tempo, nous apparaît néanmoins indispensable pour voir jusqu'où on peut pousser le *suivi de tempo* dans ce cadre minimaliste et élémentaire. Nous soulignons également par cette démarche le désir de ne pas considérer pour le moment les éléments culturels ou de hiérarchie de tempo, comme les accents qui impliquent un découpage métré, une organisation du tempo par niveaux, ou dessinent une structure plus globale de la forme musicale. S'attacher au minimum d'informations pour déduire le tempo nous permettra par la suite de considérer les autres dimensions et de trouver une façon de les prendre en compte, de les combiner dans le contexte de calcul probabiliste pour affirmer ou infirmer les inférences. Avec la seule dimension temporelle considérée, la continuité dynamique de l'inférence peut néanmoins dans de nombreux cas engendrer un suivi de tempo robuste, bien que dépendant du contexte dans lequel le suivi est propagé.

Suivi de tempo et contexte musical

Le suivi de tempo proposé ne se veut pas uniquement un suivi même si le suivi et la prédiction de la pulsation doivent être réalisés du mieux possible en fonction des observations. Il nous paraît important de rappeler que notre intention est de proposer une pulsation qui soit en interaction optimale, en *accord* avec les événements musicaux produits. À travers le choix du cadre probabiliste mis en oeuvre dans la résolution – les *indices de coïncidence* – et dans l'inférence – modèle dynamique *HMM*, nous cherchons à maximiser une fonctionnalité musicale de la synchronisation rythmique. Nous pensons que cette maximisation ou optimisation est ce qui intuitivement caractérise la transcription la plus probable. Néanmoins, l'absence totale de connaissances *a priori*

ne permet pas de contraindre l'exigence sur le temps de réaction et d'adaptation du suivi de tempo comme on peut le faire dans le cas du suivi de partition pour lequel un élément comparatif de référence est donné – c.-à-d. la partition. Le *suivi de tempo* tel qu'il est mis en œuvre dans notre thèse ne peut que renforcer sa croyance intrinsèque ou diminuer sa certitude en fonction de la suite de ses observations. Nous rappelons ici les hypothèses de notre recherche :

1. nous observons un signal musical préalablement segmenté dans le temps.
2. les oppositions de durées sont suffisantes pour créer un rythme, seule la dimension du temps est retenue.
3. l'écoute de la séquence musicale induit très probablement un tempo

Notre *suivi de tempo* est issu d'une recherche dans un cadre réduit, mais par conséquent qui reste assez général. Tout en procurant des résultats déjà satisfaisants, il apparaît comme une base prometteuse pour des recherches futures.

Contributions

Dans cette thèse, nous avons présenté une nouvelle approche pour la recherche de la pulsation ou *tactus* basée sur un développement mathématique du concept original de *grille optimale* de pulsation. Nous avons étendu et explicité à partir des études de courbe de tempo de [Guerino Mazzola \[1994\]](#) la *fonction de performance* qui associe représentation symbolique et exécution musicale. Nous avons modélisé à l'aide de formules simples le passage de l'une à l'autre de ces temporalités et l'avons appliqué dans un cadre d'accélération de tempo. Nous avons prolongé cette relation de temporalité duale dans la notion de transcription. Nous avons également modifié l'utilisation d'un algorithme basé sur la *pulsation attentionnelle* d'un oscillateur non linéaire [[Large and Palmer, 2002](#)]. Dans le cadre d'une collaboration avec l'université de Keio, nous intégrons notre suivi de tempo pour la génération de musique en interaction avec un processus visuel provenant de l'analyse d'un paysage de lumières clignotantes. Un article publié correspondant à ce travail en cours est ajouté à l'annexe (D).

6.2 Recherches futures et extensions

Contexte culturel et modes de représentation

Nous savons bien que le style musical conditionne en partie la métrique, la forme. Les éléments culturels englobent aussi le style, si bien que de nombreux paramètres peuvent être regroupés pour définir à partir de leurs valeurs un genre musical qui pourra orienter les évolutions ou choix de tempo. Ainsi dans certains styles de musique des informations extérieures de nature différente – visuelles par exemple – viennent parfois renforcer les évolutions ou les changements brusques de tempo. Une information sur la structure de la forme cyclique ou de pattern de rythmes permettra sans doute de mieux synchroniser les phrases musicales. Nous envisageons dans la suite de notre recherche d’ajouter une forme de représentation de ces connaissances lorsque le suivi de tempo est utilisé dans un style musical précis, ou comportant certains événements ou patterns particuliers. Ces paramètres sont à prendre en considération comme des éléments de partition, des indices qui peuvent augmenter une probabilité de choix à un instant donné. Nous pensons par la suite accroître le cadre probabiliste utilisé, dans une continuité des *réseaux dynamiques bayésiens* – dont la résolution par *HMM* fait déjà partie.

Calcul des inférences pour le suivi de tempo

Nous explorerons également d’autres méthodes de programmation dynamique concernant la résolution de l’inférence et de la prédiction de la pulsation. Plusieurs méthodes sont issues des techniques de *Machine Learning*, citons principalement :

- Filtre de Kalman et filtre de Kalman étendu [Maybeck, 1979]
- Chaîne de Markov Monte-Carlo (ou filtrage particulaire)[Cemgil, 2001][Cemgil, 2003]
- Inférence dans un contexte bayésien dynamiqueMurphy [2002]

Le filtrage de Kalman a pour particularité de proposer une résolution du tempo et de sa phase en même temps, il serait alors intéressant de le comparer à la reconstruction de phase de [Large and Jones, 1999] afin de regarder comment cette dernière méthode se compare au filtrage de Kalman.

Applications

Analyse Musicale Bien que le cadre principal d'utilisation de notre *suivi de tempo* soit pensé pour le temps réel, une implémentation est en cours dans l'environnement *OpenMusic* afin de proposer dans cet environnement compositionnel des analyses rythmiques de séquences musicales. Les données de tempo et de pulsation pourront alors être utilisées au gré de l'imagination des musiciens et compositeurs comme le permet ce logiciel.

Quantification et transcription Une information fournie par le décodage du tempo est la quantification. Elle renseigne de manière évidente la transcription musicale : celle-ci peut être utilisée pour des flux de données très divers provenant par exemple d'algorithmes mis au point par l'utilisateur, d'analyses spectrales, de courbes, de dessins, etc... Le problème de la quantification reste toujours ouvert dans *OpenMusic* et notre approche peut contribuer à relancer les travaux autour d'un nouveau quantificateur temporel.

Improvisation interactive *OMax*¹ est un outil d'improvisation en temps réel qui apprend des éléments de phrasé et de construction musicale à partir d'une écoute d'un flux audio d'un musicien. *OMax* utilise un algorithme de facteur d'oracle pour générer ses improvisations. La mise en place rythmique n'est pour le moment pas synchronisée avec le flux audio écouté. Notre algorithme de *suivi de tempo* devrait être à court terme intégré en temps réel à *OMax* afin de contrôler également la dimension temporelle de l'improvisation.

Suivi de tempo polyrythmique Dans le cas de sources sonores séparées – issues de captation d'instruments séparés d'une batterie par exemple –, il serait intéressant de traiter de manière séparée les analyses des flux temporels et de définir un agent de collaboration des analyses. Cela permettrait aussi de réduire le coût de calcul, la prédiction d'une pulsation pourrait être incluse dans le calcul du second flux, présentant une forme collaborative des flux temporels parallèles.

¹<http://recherche.ircam.fr/equipes/repmus/OMax/>

Validation et boucle de stabilisation

Notre suivi de tempo est construit sur des hypothèses psychocognitives et physiologiques de perception du tempo. Nous devons maintenant continuer les validations dans des situations plus complexes sur des corpus musicaux précis afin d'améliorer l'inférence et la compréhension du suivi de tempo. L'ergonomie de contrôle des paramètres est aussi un facteur déterminant pour une utilisation correcte dans les situations de performance auxquelles nous nous consacrons particulièrement. Nous nous basons alors sur le principe de stabilisation de boucle afin d'accroître les performances de notre suivi de tempo.

Annexe A

Interviews

Entretien avec Stéphane Galland¹

GN Qu'est-ce que le rythme ?

SG C'est une question tellement vaste ! Je dirais que c'est une articulation, c'est un peu comme les consonnes d'un langage, c'est ce qui rend les éléments dans le temps. Il y a aussi plein de sortes de rythmes. Ce sont différents niveaux d'articulation, comme une image qui a de plus en plus de définition. C'est une couleur, ou des couleurs qui se mélangent et ça peut devenir une forme ou un détail, pour moi, le rythme est ce qui crée cette différence, ce qui délimite des événements. Par rapport à la notion d'articulation, les niveaux d'articulation peuvent être infinis, c'est très difficile à définir par rapport à ça, car on peut avoir un son très long qui a un rythme, le rythme est présent quand quelque chose se définit dans le temps. Les conceptions de "il y a du rythme" ou de "il n'y a pas de rythme" parlent complètement d'autre chose pour moi.

GN Que représente le tempo dans ce contexte ?

¹Stéphane Galland est batteur du groupe Aka Moon. Musicien reconnu pour sa créativité, il est un des grands spécialistes des rythmes complexes, http://en.wikipedia.org/wiki/Stephane_Galland

SG Un rythme n'implique pas forcément un tempo. La frontière entre l'abstraction et le concret n'est pas vraiment nette, quand est-ce qu'il y a un tempo, et quand est-ce qu'il n'y en a pas ? Avec toutes les expériences qu'il y a eu dans la musique du *XX^e* siècle, on a été tellement loin dans l'abstraction que ça peut devenir difficile de définir un tempo. Le tempo a à voir avec un cycle, avec un mouvement, avec une perception intérieure, je ne sais pas trop comment on pourrait la définir. Mais pour moi, le tempo aurait plus à voir avec cette sorte de perception de cycle, et cela ne venant pas uniquement de la musique, c'est aussi plutôt arbitraire comme choix de définition de tempo.

GN Qu'est-ce qui fait que deux musiciens arrivent à jouer ensemble en s'écoutant ?

SG Je pense qu'il y a une sorte de régularité dans le jeu qui s'établit au cours du temps, qui établit une sorte d'accord sous-jacent entre les musiciens, qui sont ressentis de part et d'autre. Pour parler sur des exemples concrets, quand je joue avec d'autres musiciens, en fonction d'eux, il y aura une évidence de tempo plus ou moins marquée dans le jeu, quand je joue avec Nelson², la base est parfois à peine suggérée, parfois presque invisible, car notre connexion musicale est très claire, et pour cela on est ensemble. On comprend un développement rythmique, une logique dans les jeux de tension par rapport à un rythme de base. Il y a un élément qui fait référence et qui clairement définirait le tempo, dans ce cas-là c'est bien ça qui définit le tempo. Au départ, on sait quelle est la base, quelles sont les pulsations, quand je parle de ça je parle de quelque chose qui est fixé à l'avance, sur lequel on s'est mis d'accord auparavant. Sans cette donnée-là, cela serait beaucoup moins évident. Si ce n'est pas défini ou exposé avant à un moment ou un autre, je pense que ça peut être impossible de trouver le tempo de ce qu'on joue, ou alors on part dans une interprétation qui peut être complètement différente, parallèle, mais pas nécessairement connectée. Il y a parfois des introductions de morceau que Nelson joue seul où je n'arrive pas à trouver quel est son tempo, s'il démarre à un moment où je ne suis pas en phase ou pas prêt, si rien d'évident n'est marqué, alors je vais chercher parmi tous les repères possibles : rythmiques, mélodiques harmoniques, parfois c'est juste la couleur harmonique qui me permet de me repérer. Ce qui est intéressant c'est que l'harmonie est peut-être ce qu'il y a de plus flou du point de vue de la définition de tempo, par rapport à la mélodie et même au rythme, mais c'est elle qui paradoxalement permet ici de trouver

²le guitariste brésilien Nelson Veras

le tempo, certainement parce qu'elle suggère le mieux la forme. si les éléments les plus clairs ne me renseignent pas, je vais aller vers l'harmonie, pour chercher l'enchaînement des accords, et si je trouve où on est harmoniquement, peut-être que mélodiquement quelques notes préciseront ma pensée, puis l'élément rythmique sera mieux compris, par rapport au jeu que je connais de Nelson, de sa manière de développer la musique. Ce qui me permet de saisir le tempo peut être un débit de notes ou un rapport entre débits de notes, s'il commence par des triolets groupés par quatre qui peuvent créer une polyrythmie que je prends pour des croches ou des doubles croches, je peux être perdu pendant longtemps, mais le fait qu'il joue à un moment des noires ou des croches peut rendre la lisibilité du tempo évidente.

Tout cela s'inscrit bien sûr dans une forme, par rapport à des éléments qu'on connaît déjà. La déduction de notre tempo pour une personne qui écoute notre jeu est possible et dépend de la complexité polyrythmique. Personnellement, j'aime bien inclure dans mon jeu des éléments qui donnent des clefs, des indices comme des accentuations sur les points principaux, les temps, les répétitions, c'est comme donner plus de contraste pour mieux percevoir une photo.

GN Que se passe-t-il si on ne joue pas les accents ?

SG La perception est alors moins définie, comme un effet de flou, c'est moins marqué, moins *groovy*, c'est un choix. J'adore quand les choses sont tellement claires pour qu'on puisse jouer des variations extrêmes, en ayant toujours un point de repère. Je préfère ça un jeu où plus personne n'est connecté quand le jeu devient trop imprécis, car il y a une perte de racines.

GN Cette connexion, en quoi consiste-t-elle ?

SG Est-ce qu'on a les mêmes points de repère ? Pas forcément... mais oui, il faut qu'il existe des points communs, peut-être pas tous. Je me base principalement sur la ligne de basse, si elle est écrite ou si c'est dans une suite harmonique, je me base sur la fondamentale. Si elle n'est pas jouée et que je sens qu'il y a une perte d'adhérence, alors je vais la souligner. Pour rester connecté, il faut qu'il y ait une adhérence, quand elle disparaît momentanément, cela crée un flou, une tension, c'est à ce moment-là

qu'il faut être très vigilant et remettre des éléments de soutien pour éviter qu'il y ait a lors une dérive.

GN La perte d'adhérence, c'est le tempo qui bouge ?

SG Oui, mais ça peut-être aussi une interprétation incorrecte. Ça arrive très souvent avec Aka Moon que je commence à jouer un rythme très évident, mais à contre-temps, l'effet est tellement fort à cause de l'habitude qu'il peut perturber Fabrizio³ dans la construction de son développement musical. Un rythme tellement marqué peut l'entraîner à son insu, et s'il n'est pas attentif à cette illusion rythmique il ne va pas se rendre compte qu'il joue à l'envers en réalité. Si je joue tout le temps en double croche et dans un *break* je joue un quintolet, si on n'est pas habitué, on peut prendre le quintolet pour une accélération, c'est une erreur d'interprétation.

GN Quelle est la nature de cette différence, entre accélérer le jeu, et passer d'un débit de doubles-croches à un quintolet ?

SG Ça dépend du degré de perception. C'est lié à l'éducation musicale, si on n'a jamais joué un quintolet, ou un triolet, quand on l'entend on ne peut pas l'interpréter comme un quintolet, car il n'est pas dans le vocabulaire. Il y a beaucoup de groupes, indépendamment de leurs qualités musicales, qui utilisent un vocabulaire rythmique assez basique qui ne dépasse pas le triolet, même si cela n'empêche en rien une interprétation rythmique. J'ai aussi joué dans des groupes où je sais que personne n'utilise de quintolets par exemple, mais où je sens que le jeu n'est pas déstabilisé si je joue des formes rythmiques complexes. C'est aussi en raison d'une certaine confiance, quand le tempo est bien installé et bien régulier, un quintolet sonne comme un élément totalement différent. Si le tempo fluctue de manière trop forte et n'est pas assez stable, alors un quintolet devient peu évident dans sa prégnance d'où une confusion avec une variation de tempo non désirée, ce qui augmente le flou rythmique. Ça dépend de la manière dont le tempo fluctue. Quand je joue avec d'autres musiciens, j'analyse intuitivement dans quel degré de subdivision ils se trouvent, je peux voir avec qui je peux me lancer dans une complexité rythmique plus forte, avec ou sans décalages. L'idée est

³Fabrizio Cassol, saxophoniste du groupe Aka Moon

de nourrir les échanges. Quand on est décalé, c'est toujours délicat de savoir comment se recalculer, suivant qui a compris ou non qu'il y a un décalage, alors si le flou total arrive il faut des indices forts pour resynchroniser tout le monde.

GN Quelle sont les qualités du tempo ?

SG Je pense que le tempo est comme une grille, comme une base solide, un sol, sur lequel tout se construit, c'est une référence. S'il y a variation de tempo, elle doit se faire avec une certaine douceur, s'il y a un angle ce n'est pas une variation de tempo. Les variations de débit sont des variations plutôt carrées, car plus radicale. Si le tempo change radicalement alors on peut l'interpréter comme un changement de débit. Bien sûr, il y a beaucoup de partitions où le tempo change brusquement, mais pendant ce changement-là, souvent les composantes harmoniques ou mélodiques accompagnent ce changement, donc on perçoit un changement de tempo. Si tout reste pareil sauf le tempo, cela peut être senti comme un changement de tempo, mais est plus probablement un changement de débit.

GN Cela devient alors une polyrythmie ?

SG Oui, c'est ça, une polyrythmie.

GN Comment bouge le tempo quand vous jouez ensemble ?

SG Si je veux créer une accélération ou une décélération, je vais essayer d'être le plus doux possible pour que ce ne soit pas perçu comme un changement de débit. En ce sens, cela sera différent d'un jeu où je suis en triple croche et je passe en septolet, sextolet et quintolet, cela serait trop net pour être un ralentissement. Un ralentissement doit être courbe, arrondi. Un changement de tempo pourrait être aussi radical et être perçu tel quel si les autres éléments sont connectés à ce changement. Si la ligne mélodique et harmonique varient aussi de la même manière, et surtout si elles ont été exposées avant, on va mieux percevoir cette accélération brusque. Mais ça peut être aussi une variation, comme pour la musique indienne où on a beaucoup d'illusion de tempo, ce

qui rend la chose assez complexe : quand est-on dans une illusion de tempo et quand est-ce que le tempo change réellement ? En général, à un certain niveau de jeu, on ne fait pas de variations énormes de tempo, ça vient subtilement et ça bouge d'une manière assez fluide. Si cette fluidité est absente, on peut alors se poser la question d'un changement de tempo ou d'une polyrythmie voulue. Ça arrive aussi qu'une phrase clef réalise un changement de tempo. Il s'agit bien sûr d'une phrase qu'on connaît, qui fait démarrer un morceau, si à un moment quelconque je rejoue cette phrase à un tempo différent, alors tout le monde comprend qu'on vient de changer de tempo, parce que cette phrase est préméditée. IL y a donc une connaissance musicale qui intervient dans ce cas.

GN Quelle conscience du mouvement de tempo as-tu ?

SG Étonnamment, on ne se rend pas toujours compte si le tempo bouge. Parfois ça ne sert à rien de garder un tempo fixe, car il y a une tendance plus générale contre laquelle on ne peut pas aller. Cela m'est arrivé avec un groupe où on jouait avec des machines, ma partie de batterie était doublée avec des parties programmées et je ne pouvais pas bouger et, surtout au début de la tournée, tout le monde accélérail. C'était une sensation très désagréable de ne pas pouvoir les accompagner dans cette accélération. La batterie paraît alors très lourde, surtout pour moi qui aime bien jouer un peu devant, c'était vraiment à l'opposé de mon jeu. Jouer exactement sur le temps ne sonne pas pour moi, je l'ai expérimenté, mais je trouve que ça casse le mouvement, la danse, il n'y a plus cette fluctuation qui donne le mouvement, pour cela le rapport entre les musiciens est très important. Le placement rythmique des musiciens crée une profondeur. C'est la même chose au piano pour certains musiciens, un décalage même très faible entre les notes d'un accord permet d'ouvrir la sonorité, les différentes hauteurs ont aussi des inerties différentes et donc des retards inhérents dans leur émission sonore.

GN La distance entre les musiciens induit des retards du même ordre également.

SG Oui, ça crée une richesse comparable aux harmoniques d'un son. Mais quand je joue avec un bassiste qui aime aussi jouer devant le temps, c'est plus difficile. Je dois faire des efforts pour rester légèrement en arrière, sinon on avance tous les deux

et on peut très facilement accélérer, comme avec Nelson, par exemple qui aime aussi généralement jouer en avant du temps ! Avec Michel⁴ il y a un équilibre qui fonctionne très bien, il joue profondément dans le temps, peut-être que s'il joue avec un batteur très *laid-back* le tempo risque de tirer en arrière et il devra modifier son jeu. Ce sont des questions de perception, de goût, de personnalité. Dans la batterie même je vais jouer la caisse claire plus devant que la grosse caisse. Certains éléments dynamisent le jeu, d'autres l'assoient comme la grosse caisse, sont plus réguliers comme le *HiHat*. Fabrizio et Michel demandent toujours du HiHat dans les retours, car ils se basent dessus, je l'utilise aussi pour pouvoir créer des tensions plus fortes grâce à cet élément stable. IL y a toujours un élément qui montre où on est. On construit une tension par rapport à un point. Si ce point n'est plus là, cette forte tension peut sonner comme un accord parfait alors qu'il est basé sur deux autres notes qui créent cette tension harmonique importante. C'est pour ça que j'aime bien que la basse soit soulignée subtilement, qu'elle soit perçue.

GN Est-ce qu'il existe un partage de la pulsation ?

SG Oui, tout à fait, il y a clairement une pulsation dans notre musique. Il y a souvent des gens qui dansent en écoutant Aka Moon, malgré nos changements de mesure ! Plus on crée des tensions, plus on s'éloigne et moins cette pulsation est évidente. Ma perception est qu'une fois le tempo installé, on peut créer toute sorte de tensions, s'éloigner même du tempo, mais tant que je perçois intérieurement cette pulsation, ce tempo, je ne vais pas redevenir plus évident rythmiquement. Si je sens dans ma propre perception ou dans celle des autres, que la pulsation disparaît, alors je vais suspendre mon développement et matérialiser de façon plus évidente le tempo. Parfois, même le public peut avoir une influence, si je vois que mes idées rythmiques cassent leur perception, alors je peux jouer par rapport à ça, bien sûr c'est plus un sentiment, une intuition, mais on saisit quand même cette attention chez les personnes qui écoutent.

GN Parfois tu ne joue pas cette pulsation, y-a-t-il un rapport avec certaines musiques ethniques ?

⁴Michel Hatzigeorgiou, bassiste du groupe Aka Moon

SG Oui, c'est incroyable les musiques ethniques ! Au début j'ai eu beaucoup de mal parce qu'ils mettent des accents complètement inhabituels pour nous. Par exemple pour les musiques africaines en général, qui sont ternaires, ils mettent souvent l'accent sur le second triolet, et c'est un accent vraiment très fort, il ne montre pas le temps comme nous le pensons. Pour eux, le temps est presque ce qui est le moins marqué. C'est la base, c'est ce qui est évident, les pas de danse sont clairement sur les temps, mais tout le reste va en balance avec ça. Il y a des rythmes du Maghreb que j'ai appris, des rythmes traditionnels où les accents ne sont pas du tout sur le premier temps. Au début je ne comprenais pas, car ça sonne tellement évident avec la basse sur le premier temps. Maintenant, je me suis habitué à ce sentiment de ne pas marquer le premier temps. Pour repérer ce premier temps, c'est souvent le chant qui commence dessus. Ce sont des musiques avec des cycles qui se répètent et où le premier temps n'est jamais marqué. Parfois arrivent soudainement des claquements de mains, ils sont alors toujours sur les temps forts ! Quand on s'habitue à ce son-là, on sent quelque chose de complet dans cette façon de jouer : le premier temps, la base, est là, elle est ancrée, il n'y a alors pas besoin de le souligner. Tout ce qu'ils font de générer des tensions en balance avec ce temps et cette pulsation. Il y a une forte présence et sensation de danse dans ces musiques, les tensions sont en équilibre. Les accents sont comme en contradiction avec ce qu'ils permettent de sentir dans le mouvement par exemple.

GN C'est différent en ce sens de la musique indienne.

SG La musique indienne est beaucoup plus narrative. En Inde du Nord, le repère de la pulsation est plus auditif, en Inde du Sud, il est plus visuel, par rapport au compte des temps : ils savent à quel moment, à quelle double croche de quel temps ils peuvent commencer une phrase rythmique qui se finira sur le *sam*⁵ après avoir été répétée trois fois. Ce sont des portes qui s'ouvrent à un endroit et se ferment à un autre, c'est très intéressant, ils ouvrent des perceptions et des temps différents. Dans un autre domaine, un pianiste comme Pogorelich crée des sensations identiques. Il dérange beaucoup de personnes en raison de sa manière de jouer très personnelle. Je me suis rendu compte qu'il ne donne pas une perception du temps linéaire dans sa façon de jouer. C'est une perception du temps à des niveaux différents. Si on cherche une perception linéaire dans l'écoute de son jeu, cela peut être très désagréable, car on ne comprend pas le sens rythmique dans un temps linéaire, dans un autre temps

⁵le premier temps d'un cycle rythmique en musique indienne

cela devient fabuleux. C'est un temps fonction d'autres paramètres, de plans sonores, c'est difficile à expliquer, mais il y a comme une relation entre le son et comment il se détache du temps. Si on ne les perçoit pas de cette façon, on ne comprend pas pourquoi dans le temps linéaire il y a une coupure par exemple, c'est comme s'il y avait plusieurs temps parallèles dans lesquels il se déplace en fonction des sonorités. On sent ça moins sur les enregistrements, mais en concert c'est assez évident, dans ses interprétations de Rakmaninov, des sonates de Beethoven aussi qui peuvent sonner comme du Debussy ou de la musique contemporaine alors qu'il joue le texte, mais ses accentuations emmène la musique ailleurs, c'est très surprenant.

Il y a des musiciens qui ont plutôt un tempo rapide et d'autres plutôt un tempo lent. Si on analyse Tony Williams⁶ et Elvin Jones⁷, ou bien Charlie Parker et John Coltrane, c'est très évident que Tony Williams a un tempo rapide, si on l'analyse temps par temps c'est assez clair, par contre analyser Elvin Jones temps par temps est très difficile, voire impossible ! Par contre dans un tempo lent, on comprend que sa perception est dans les grands cycles, alors tout devient très clair, il se base sur des accents lents, à l'intérieur desquels il a des phrasés très courbes. Ses accents et sa manière de penser un tempo lent rendent la compréhension plus évidente. De même pour Charlie Parker, si on analyse son jeu à la noire, on peut tout transcrire facilement. Si on écoute Coltrane, le débit est complètement différent, on sent les courbes, où elles vont, où elles commencent, où elles atterrissent. Il s'agit d'une perception plus longue, même si à l'intérieur de cette lenteur cela peut être beaucoup plus complexe.

GN Il s'agit-là d'une échelle de temps différente.

SG Oui, j'ai l'impression. La pulsation se fait plus évidente dans une échelle de temps plus grande ou bien plus resserrée. Je pense que ça doit être plus facile de faire un suivi de tempo sur Tony Williams que sur Elvin Jones. Mais si on envisage des fluctuations avec des repères à long terme assez réguliers, alors ça sera peut-être plus facile sur Elvin Jones.

GN Cela pose pour l'ordinateur le problème de perception des éléments qui induisent la pulsation. Je t'ai parlé d'un algorithme qui recherche une grille régulière de pulsa-

⁶Tony Williams (1945-1997†), batteur de Miles Davis de 1963 à 1969

⁷Elvin Jones (1927-2004†), batteur de John Coltrane de 1960 à 1966

tions par rapport à des événements dans le temps. Est-ce qu'il t'arrive de jouer en pensant à une pulsation que tu ne joues pas du tout, qui n'existe plus que dans la pensée, mais sans plus aucun élément susceptible de la suggérer dans ton jeu ?

SG Je pense que ça ne m'est jamais arrivé. Ou bien à l'époque où je jouais plus de *free*. Parfois avec Nelson en duo, cela va très loin dans la suggestion de tempo. On joue quand même toujours par rapport à une structure qui est claire. Il y a toujours des moments où on va se retrouver, même si c'est très peu. Notre référence est vraiment la structure, si elle disparaît, cela devient complètement autre chose, qui peut certainement être intéressant, mais c'est par rapport à une base qu'on construit nos improvisations et qu'on cherche à se retrouver. C'est comme une peinture ou un dessin abstraits qui partent originellement d'une figure. On pousse l'abstraction le plus loin possible, mais il y a encore une idée de la forme figurative, réduite à sa plus simple expression. Picasso décline plusieurs peintures d'un taureau de cette manière. Il arrive à un résultat très dépouillé, très simple et très peu réel. La première peinture est très évidente, un taureau assez classique. Par des développements successifs, il transforme le taureau jusqu'à le réduire à quelques lignes qui étonnamment le suggèrent toujours de manière frappante, même si ce résultat est vraiment très éloigné de l'original. On joue parfois *Besame mucho* en septolets, et Nelson ne rend pas du tout les harmonies évidentes. Tout le monde connaît *Besame mucho*, mais même des musiciens ne reconnaissent pas, il reste une certaine qualité ou saveur de ce morceau et je pense que c'est ça qui touche les gens même s'ils ne reconnaissent pas la musique. Quand je vois des peintures, il y a quelque chose que je ne comprends pas toujours, mais que je ressens. Je pense que musicalement c'est du même ordre, il y a des réminiscences qui nous envahissent.

GN Oui, même si la perception de la forme est dans une dimension différente, globale et spatiale pour la peinture, et développée dans le temps pour la musique. Il y a une mémoire d'éléments dans l'espace et une mémoire d'éléments dans le temps. Le jeu musical combine à lui seul ces deux aspects de l'espace et du temps, les mouvements dans leurs moindres détails créent des coupures du temps, accentuées en fonction des déplacements et forces mises en jeu.

SG C'est vrai, et les dynamiques sont aussi très importantes dans la perception du rythme et donc du tempo. Même si ce n'est pas impossible, c'est plus complexe de comprendre la construction du rythme sans tenir compte des accents. Les hauteurs de son sont également une information importante pour comprendre le rythme, on sent qu'il y a un temps qui tombe et un temps qui monte, même s'ils sont joués comme des contre-points dans les rythmes africains. Il y a des personnes qui comprennent très rapidement et intuitivement une structure rythmique, et d'autre comme moi qui doivent réfléchir pour être sûr de bien la saisir ! Sans doute à cause de notre éducation qui nous a appris à noter la musique.

GN Qu'est-ce que tu attendrais d'un ordinateur qui joue avec toi ?

SG Suivre les petites inflexions qui sont toujours présentes dans le jeu. Déduire, aller dans une imagination folle, si la machine peut apprendre aussi. Le chemin vers une interprétation plus abstraite vient d'un travail très clair initialement, cela serait intéressant de pouvoir faire ce chemin avec la machine, qu'elle puisse apprendre certaines subtilités de jeu au fur et à mesure, qu'elle se crée ses propres points de référence. Quand on joue ensemble, on cherche à être connectés, l'analyse est donc très rapide localement, mais aussi à un niveau plus lent par rapport au temps de la métrique. Le jeu rythmique laisse souvent des moments indécis qui sont clarifiés rétrospectivement par la suite, c'est ce qui doit être difficile pour l'ordinateur je suppose.

Entretien avec Riccardo Del Fra⁸

GN Je voudrais qu'on parle du temps musical et de son rapport au rythme, au tempo.

RDF Est-ce qu'on parle d'un temps individuel ou bien est-ce qu'on pense dans un partage espace-temps ? Je pense que c'est légèrement différent. Si je fais un solo, naturellement je serais plus apte à me libérer d'un métronome dans la tête ou réel. Si je joue avec quelqu'un, d'un duo à un grand orchestre avec un chef, il y a un partage qui fait que la définition du temps change.

GN Oui, c'est bien de ce partage dont il est question. Mais il peut s'agir de ton temps individuel afin qu'il puisse être compris par un autre musicien ou même un ordinateur. Comment construis-tu le temps musical ?

RDF Je considérerais que je suis seul premièrement. En tant qu'improvisateur, je décide de jouer quelques miniatures par exemple, comme des haïkus où chacun aura sa personnalité. Je me donne des contraintes, la première miniature pourra être construite autour d'intervalles très rapprochés, des secondes mineures ou majeures, voire au maximum un intervalle de tierce majeure. Les contraintes peuvent être les nuances, et donc relatives à la dynamique. Je peux commencer à jouer *rubato*, ou bien avec un tempo très lent, mais que je peux interpréter de façon très libre comme un musicien classique. Tu me connais, je considère aussi beaucoup le silence, c'est une part importante de la chose musicale. Si le premier haïku est *rubato*, le second pourra être assez rythmé dans lequel j'utilise une série d'intervalles plus larges dans un registre médium aigu pour créer un espace différent du précédent, et dans le rythme j'essaie de penser exclusivement binaire par exemple. Là, je me crée un partenaire idéal, éventuellement celui qui écoute, et j'essaie d'être le plus précis et rigoureux possible. Si je pense binaire, toutes les approches sur les notes courtes seront de l'ordre de la croche binaire, de la double-croche, etc... et sans triolet si je suis strict sur mon concept. Pour le troisième, je pourrais avoir une approche ternaire, et si mon débit est irrégulier, le phrasé sera

⁸Riccardo Del Fra est contrebassiste et compositeur, il a joué notamment avec Chet Becker de 1979 à 1988. Riccardo Del Fra est responsable du département Jazz et Musiques Improvisées du CNSMDP depuis 2004

absolument ternaire, même les noires ne seront pas les mêmes que les noires binaires précédentes. Je pense que ce qui est sous-jacent me prépare à une forme de souplesse ou pas. Évidemment, il faut être souple dans les deux cas, mais on est plus *carré* au niveau de la diction quand on est binaire, dans le triolet il y a quelque chose de plus *rond*, je sais qu'on touche ici quelque chose qui n'est pas forcément rationnel.

GN Cette perception est très subtile effectivement.

RDF Elle rejoint peut-être certaines réflexions algorithmiques pour la perception par l'ordinateur du geste musical. Cela serait fantastique si l'ordinateur pouvait distinguer un jeu binaire d'un jeu ternaire, et l'utiliser comme un paramètre pour sa réponse ou la transformation du son. Il pourrait rapidement reconnaître si on pense binaire ou ternaire.

GN Dans ton analyse rythmique, d'où vient cette pensée binaire ou ternaire ?

RDF Si je continue dans ma réflexion sur la construction des improvisations, pour contraster je pourrais introduire dans la miniature suivante des éléments bruités comme frotter la touche donner un coup sur le chevalet, ou taper une corde bloquée sans créer de hauteur précise. Là aussi, ce sont des gestes rythmiques, sauf qu'ils sont dans une sorte de "désordre" par rapport à là d'où l'on vient. Si le cinquième haïku est un climax, je reprendrai les idées rythmiques binaires et ternaires mélangées, sur un tempo plus rapide, avec toute sorte d'intervalles pour revenir finalement à un tempo apaisé dans le sixième et dernier haïku. Dans la phase de climax, l'idéal est le mélange de binaire et de ternaire. Je prends un exemple avec Marc Johnson⁹, quand il joue avec Enrico Pieranunzi¹⁰ ou dans un contexte plus postmoderne, on entend exactement cela. Donc, les deux aspects peuvent cohabiter, même si à un moment donné on ne sait plus quel est celui qui prédomine. C'est la combinaison des éléments qui est créative. Chet Baker joue par exemple sur une octave et demie toujours le même genre de phrases, mais il les combine de façon incroyablement créative, on entend souvent qu'il est en train de chercher, et le silence est d'une importance capitale. Chet s'amuse à finir sa phrase

⁹ compositeur et contrebassiste américain

¹⁰ pianiste de jazz italien

non pas en pensant tellement à la phrase, mais en cherchant où il va placer le silence. Tout cela pour moi fait partie du *time*, du tempo.

GN Le *time* et le tempo, c'est la même chose ?

RDF Oui, absolument, c'est la même chose. Quand je joue seul, j'ai des libertés que je ne peux pas avoir avec quelqu'un d'autre. Le tempo, quand on est en duo, ce n'est ni moi ni l'autre, c'est à la limite ce qui est entre nous deux, ce qui appartient à tous les deux, ce qui est là. Quand on entend un duo, un bassiste et un pianiste qui jouent un peu *straitgh ahead* sur un standard, si le pianiste accélère un peu dans sa fougue et que le bassiste reste sur son tempo sans bouger, selon moi, il n'y a déjà plus de musique, ou bien il y a un effet musical discutable, car le partage du tempo est trop relatif. La musique c'est d'être ensemble et de construire ensemble au fur et à mesure, dans cet espace-temps qu'on partage. Sinon on est dans toute autre chose, lorsqu'on joue avec un clic pour un enregistrement, je le fais pour le cinéma, car il y a une synchronisation avec l'image précisément, mais c'est très différent du point de vue de la musique, c'est très rigide.

GN Comment le tempo bouge-t-il dans le jeu ?

RDF IL y a une élasticité, il y a des courbes à l'intérieur, qui sont inimitables presque. Quand on écoute l'enregistrement célèbre *Four and more* de Miles Davis, les morceaux commencent à un tempo bien inférieur à celui auquel ils se terminent. Ce n'est pas parce qu'il le fait que tout le monde doit faire pareil. Néanmoins, pour un morceau *bop*, si ça ralenti, c'est fatal, on ne peut pas, le tempo ne peut qu'accélérer, car il y a un sens musical, on va vers un climax, et le contraire est à mon avis dangereux. Pour d'autres morceaux, c'est intéressant d'entendre le dernier thème qui s'effiloche, on a alors un *rallentando* qui dure un certain temps, c'est un effet très beau aussi. Les enregistrements ECM¹¹ offrent une autre attitude pour juger de ce genre de chose que le contexte *bop*. Cela dépend du style, du contexte esthétique de la musique.

¹¹ECM est un label de musique réputé pour la qualité et l'originalité de ses enregistrements, où le sens de l'espace et la clarté du son sont particulièrement mis en valeur.

GN En quoi la pulsation et l'harmonie sont-elles liées dans la compréhension du tempo ? Peut-on faire abstraction de certains éléments et toujours saisir le tempo ?

RDF Est-ce qu'on peut conserver seulement les impulsions pour percevoir la pulsation ? Oui, je pense, absolument. Si on joue plus ou moins en place par rapport à cette pulsation, on doit la ressentir. Je peux avoir en tête un débit de croches très rapide. Une appoggiature, une petite *ghost note* peuvent être aussi très importantes pour définir le tempo. Je ne sais pas si concernant le *time*, je me place de la même manière pour la même note, de même pour le son. Je pense que tout se définit au fur et à mesure ensemble, avec les autres musiciens. Un batteur peut réagir comme il veut sur les messages que je lui envoie, il peut prendre la liberté d'être binaire, ternaire, 12/8, 3, 6 à la limite il peut être en groupe de 5 etc... mais à un moment donné on reviendra, avec une convention tacite même, au tempo du début. C'est-à-dire qu'il y a quelque chose de plus ou moins stable malgré les transformations traversées. C'est une convention tacite en jazz de se dire qu'on part de quelque chose, et qu'on y revient. La rupture serait plutôt un antagonisme, et donc utilisée pour générer un fort contraste.

GN Est-ce qu'il y a un *time* absolu comme l'oreille absolue ?

RDF C'est plus ou moins relatif ! Je ne pense pas, un musicien peut savoir si on accélère ou ralentit de manière très précise. S'il y a une forme de perception rythmique absolue, je pense que c'est dans la reconnaissance des cycles. Certains musiciens peuvent reconnaître tout de suite si on vient de jouer un 11/8, ou un 23/16. Il y a une pièce que je joue avec Annie Ebrel¹², c'est un 11/8, un jour après le concert un batteur que je connais bien me dit : "c'était très bien ça, c'était en 9 ?". La perception de la structure est quelque chose qu'on apprend à force de jouer, la perception des cycles. Cela devient une grille, dans le sens grillage¹³ mais aussi un point de repère. C'est comme *I got rhythm*¹⁴, si on pense la pulsation localement, c'est difficile. C'est beaucoup plus facile de penser la structure toutes les 4, 8 ou 16 mesures. On approche peut-être les concepts de cycles de la musique indienne où c'est le cycle qui se répète dans les formes *Raga*.

¹²Annie Ebrel est une chanteuse de chant traditionnel breton

¹³car grille fait référence à la grille harmonique dans le vocabulaire jazz

¹⁴standard de Jazz de George Gershwin, aussi très connu sous le titre de *Rhythm changes*

GN Il y a quand même des moments où le partage de la pulsation est assez difficilement perceptible, non ?

RDF Oui, c'est souvent parce qu'il y a un réel plaisir de continuer dans sa lancée qui fait que les musiciens semblent avancer parallèlement. C'est comme Tony Williams avec Miles Davis, le phrasé des gestes de Wayne Shorter et de Herbie Hancock sont complètement virevolté, transformés, transfigurés par les gestes de Tony Williams : les équivalences, les dédoublements, les triolets qui arrivent tout de suite... C'est le plaisir d'être là-dessus qui fait que cela tient ensemble, il n'y a pas de déstabilisation, les musiciens continuent dans leur *time*, ils ne changent pas de direction, rien ne les y contraint, à un moment donné ils retombent ensemble. Il y aussi toujours une souplesse des musiciens dans le tempo. Quand un batteur double son tempo, il accélère parfois un peu, les autres musiciens ne restent pas stoïques sur leur tempo et adaptent leur jeu et figures rythmiques à la nouvelle pulsation, au bout de quelques temps ou mesures. C'est comme les mouvements dans la musique classique, souvent un *forte* s'accompagne d'une petite accélération de tempo.

La musique indienne présente des variations de tempo qui sont beaucoup plus fortes, mais cela est fait avec une telle science et une telle finesse qui exprime bien à quel moment cela arrive qu'on ne perd pas la forme. On suit le mouvement dans notre façon de compter, on accélère avec eux, car la phrase mélodique ou le chant porte aussi cette accélération. Je pense de plus en plus que c'est inhérent à la musique, c'est un *sine qua non* dans le partage, il y a des courbes, il y a du mouvement. Le mouvement en général est toujours là, plus ou moins évident, mais présent.

Annexe B

Elements de Probabilité

B.0.1 Elements de probabilité

Ces éléments sont issus pour la plupart de [Chung \[1978\]](#).

Dénombrement

Coefficient binomial

$$\binom{m}{n} = \frac{m!}{n!(m-n)!} = \binom{m}{m-n}$$

$$(a+b)^n = \sum_{k=0}^n \binom{m}{n} a^k b^{n-k}$$

Triangle de pascal

$$\binom{m}{n} = \sum_{k \leq m-1} \binom{k}{n-1}$$

Variables aléatoires

Définition B.0.1. On appelle variable aléatoire [sur Ω] une fonction à valeur numérique X de ω définie sur le domaine Ω :

$$\omega \in \Omega : \quad \omega \rightarrow X(\omega) \quad (\text{B.0.1})$$

Proposition B.0.1. Si X et Y sont des variables aléatoires, il en est de même pour

$$X + Y, \quad X - Y, \quad XY, \quad X/Y (Y \neq 0), \quad aX + bY \quad (a, b \in \mathbb{R}^2) \quad (\text{B.0.2})$$

Proposition B.0.2. Si φ est une fonction de deux variables (ordinaires) et X et Y sont des variables aléatoires, alors

$$\omega \rightarrow \varphi(X(\omega), Y(\omega)) \quad (\text{B.0.3})$$

est aussi une variable aléatoire, on la note $\varphi(X, Y)$

Si Ω est dénombrable, X prend ses valeurs dans l'ensemble suivant

$$V_X = \bigcup_{\omega \in \Omega} \{X(\omega)\} \quad (\text{B.0.4})$$

Introduisons la notation

$$p_n = P(X = v_n) \quad , \quad v_n \in V_X \quad (\text{B.0.5})$$

Les p_n , ou probabilités élémentaires de la variable aléatoire X , vérifient

$$(i) \quad \forall n : p_n \geq 0, \quad (\text{B.0.6})$$

$$(ii) \quad \sum_n p_n = 1 \quad (\text{B.0.7})$$

Définition B.0.2. On appelle fonction de répartition (cumulative) de X la fonction

$$F_X(x) = P(X \leq x) = \sum_{v_n \leq x} p_n \quad (\text{B.0.8})$$

Définition B.0.3. Pour une variable aléatoire X définie sur un espace Ω , son espérance

mathématique est donnée par

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) \quad (\text{B.0.9})$$

à condition que la série converge absolument,

$$\sum_{\omega \in \Omega} |X(\omega)|P(\{\omega\}) < \infty \quad (\text{B.0.10})$$

dans ce cas l'espérance mathématique existe.

(B.0.9) contient une expression pour l'espérance de toute fonction de X

$$E(\varphi(X)) = \sum_{\omega \in \Omega} \varphi(X(\omega))P(\{\omega\}) \quad (\text{B.0.11})$$

également sous condition de convergence absolue.

car de la proposition (B.0.2) il suit

$$E(\varphi(X)) = \sum_n p_n \varphi(v_n)$$

Notons que l'espérance de $\varphi(X)$ est déjà déterminée par la probabilité de distribution de X (et bien sûr aussi par la fonction φ), sans l'intervention de la probabilité de distribution de $\varphi(X)$ elle-même. Ainsi, pour $\varphi(x) = x^r$, on définit le r^{ieme} moment de X :

$$E(X^r) = \sum_n p_n v_n^r \quad (\text{B.0.12})$$

Définition B.0.4. *Considérons une fonction f définie sur $\mathbb{R}$*

$$u \rightarrow f(u)$$

et qui satisfait aux deux conditions

$$\begin{aligned} (i) \quad & \forall u : f(u) \geq 0; \\ (ii) \quad & \int_{-\infty}^{\infty} f(u) du = 1 \end{aligned}$$

une telle fonction est appelée fonction de densité sur $\mathbb{R}$

Nous pouvons définir une classe de variables aléatoires sur un espace d'échantillons indénombrable, dont la probabilité est décrite au moyen d'une fonction de densité f , telle que , si A est une union d'intervalles, nous avons

$$P(X \in A) = \int_A f(u)du$$

La fonction de répartition s'écrit alors :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(u)du \quad (\text{B.0.13})$$

et définit la primitive de f . Si f est continue, alors f est la dérivée de F

$$F'(x) = f(x)$$

On définit alors l'espérance :

$$E(X) = \int_{-\infty}^{\infty} uf(u)du \quad \text{sous la condition} \quad \int_{-\infty}^{\infty} |u|f(u)du \leq \infty \quad (\text{B.0.14})$$

Remarquons, alors que la densité continue est une idéalisation mathématique, nous avons

$$P(X = x) = \int_x^x f(u)du = 0$$

Définition B.0.5. *La variable aléatoire la plus générale est une fonction X définie sur un espace d'échantillons Ω tel que, pour tout réel x , la probabilité $P(X \leq x)$ est définie.*

Ce qui vient en premier est une *mesure de probabilité* P définie sur une *classe de sous-ensembles* de Ω . On appelle cette classe un *champs d'échantillons de Borel*, ou *champs de probabilité*, on le note $\mathcal{F}$. Si une fonction X a la propriété que pour tout x l'ensemble $\{\omega | X(\omega) \leq x\}$ appartient à la classe $\mathcal{F}$, alors X est une variable aléatoire.

Vecteurs aléatoires

Soit X et Y deux variables aléatoires à valeurs dénombrables (X prend ses valeurs

dans $\{x_i\}$ et Y prend ses valeurs dans $\{y_j\}$). Posons

$$P(X = x_i, Y = y_j) = p(x_i, y_j)$$

Quand x_i et y_j parcourent toutes les valeurs possibles, l'ensemble des probabilités élémentaires $p(x_i, y_j)$ donne la *distribution de probabilité jointe* du *vecteur aléatoire* (X, Y)

Pour obtenir la distribution de probabilité de X seul, nous laissons y_i parcourir toutes les valeurs possibles

$$P(X = x_i) = \sum_{y_j} p(x_i, y_j) = p(x_i, *)$$

Quand x_i parcourt toutes les valeurs possibles, l'ensemble des $p(x_i, *)$ donne la *distribution marginale* de X . On a de même la distribution marginale de Y avec

$$P(Y = y_j) = \sum_{x_i} p(x_i, y_j) = p(*, y_j)$$

Ces distributions marginales ne permettent pas en générale de déterminer la distribution jointe.

Nous pouvons aussi exprimer l'espérance d'une fonction de (X, Y) :

$$E(\varphi(X, Y)) = \sum_{x_i} \sum_{y_j} \varphi(x_i, y_j) p(x_i, y_j) \quad (\text{B.0.15})$$

Dans le cas de densité de probabilité, le vecteur aléatoire (X, Y) a une fonction de densité jointe f telle que pour tout (x, y)

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \quad (\text{B.0.16})$$

ou plus généralement pour un sous-ensemble *Borélien* S du plan cartésien on a :

$$P((X, Y) \in S) = \iint_S f(u, v) du dv \quad (\text{B.0.17})$$

La fonction de densité jointe f satisfait aux conditions suivantes

$$(i) \quad f(u, v) \geq 0 \quad \text{pour tout } (u, v)$$

$$(ii) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) du dv = 1$$

On obtient alors les fonctions de répartition marginales de X et de Y :

$$P(X \leq x) = \int_{-\infty}^x f(u, *) du, \quad \text{où } f(u, *) = \int_{-\infty}^{\infty} f(u, v) dv \quad (\text{B.0.18})$$

$$P(Y \leq y) = \int_{-\infty}^y f(*, v) dv, \quad \text{où } f(*, v) = \int_{-\infty}^{\infty} f(u, v) du \quad (\text{B.0.19})$$

$u \rightarrow f(u, *)$ et $v \rightarrow f(*, v)$ sont respectivement les *fonctions de densité marginale* de X et de Y . Elles sont dérivées de la fonction de densité jointe après intégration sur la variable qui n'est pas en question.

L'espérance pour un vecteur aléatoire (X, Y) avec densité de probabilité s'exprime ainsi

$$E(\varphi(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(u, v) f(u, v) du dv \quad (\text{B.0.20})$$

De façon générale, la *fonction de distribution jointe* F de (X, Y) est définie par

$$F(x, y) = P(X \leq x, Y \leq y) \quad \text{pour tout } (x, y)$$

Si on note $\lim_{y \rightarrow \infty} F(x, y)$ avec $F(x, \infty)$, on a

$$F(x, \infty) = P(X \leq x, Y \leq \infty) = P(X \leq x)$$

$x \rightarrow F(x, \infty)$ est la *fonction de distribution marginale* de X . On définit de même la fonction de distribution marginale de Y : $y \rightarrow F(\infty, y)$

Probabilité conditionnelle

La probabilité d'un ensemble A est sa proportion pondérée relativement à l'espace d'échantillons Ω . Si Ω est fini et que tous ses points échantillons ont le même poids (

i.e. la même probabilité) , alors

$$P(A) = \frac{|A|}{|\Omega|}$$

Quand Ω est dénombrable et chaque point ω a le poids $P(\omega) = P(\{\omega\})$ alors

$$P(A) = \frac{\sum_{\omega \in A} P(\omega)}{\sum_{\omega \in \Omega} P(\omega)} = \frac{\sum_{\omega \in A \cap \Omega} P(\omega)}{\sum_{\omega \in \Omega} P(\omega)} \quad \text{avec} \quad \sum_{\omega \in \Omega} P(\omega) = 1$$

En déplaçant l'attention de Ω vers un ensemble S qui devient le nouvel univers, on introduit la notion de probabilité conditionnelle

$$P(A|S) = \frac{\sum_{\omega \in AS} P(\omega)}{\sum_{\omega \in S} P(\omega)} = \frac{P(AS)}{P(S)}$$

La plupart des problèmes de probabilité mettent en oeuvre plusieurs événements ou variables aléatoires pour lesquels on étudie leurs relations mutuelles ou leurs actions jointes. Pour cela, voici quelques propositions fondamentales faisant appel aux probabilités conditionnelles :

Proposition B.0.3. *Pour des événements arbitraires $A_1, A_2, \dots, A_n$ nous avons*

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1})$$

sous la condition $P(A_n|A_1 A_2 \dots A_{n-1}) > 0$

Cette formule est appelée *formule multiplicative générale* pour la probabilité d'une intersection. Supposons n variables aléatoires $X_1, X_2, \dots, X_n$ à valeurs dénombrables. Pour des valeurs possibles arbitraires $x_1, x_2, \dots, x_n$ posons

$$A_k = \{X_k = x_k\} \quad \text{i.e.} \quad A_k = \{\omega | X_k(\omega) = x_k\} \quad , \quad k = 1, 2, \dots, n$$

nous obtenons

$$\begin{aligned} &P\{X_1 = x_1; X_2 = x_2; \dots; X_n = x_n\} \\ &= P\{X_1 = x_1\} P\{X_2 = x_2 | X_1 = x_1\} P\{X_3 = x_3 | X_1 = x_1, X_2 = x_2\} \\ &\quad \dots \quad P\{X_n = x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} \end{aligned}$$

Le premier terme ci-dessus est appelé *probabilité jointe* de $X_1, X_2, \dots, X_n$, exprimée ici par des probabilités conditionnelles successives.

Proposition B.0.4. *Supposons que*

$$\Omega = \sum_n A_n$$

est une partition de l'espace des échantillons en des ensembles disjoints. Alors, pour un ensemble B quelconque nous avons

$$P(B) = \sum_n P(A_n)P(B|A_n)$$

Cette formule est appelée formule de *probabilité totale*.

Proposition B.0.5. *Sous l'hypothèse et la notation de la proposition B.0.4, nous avons également*

$$P(A_n|B) = \frac{P(A_n)P(B|A_n)}{\sum_n P(A_n)P(B|A_n)}$$

avec la condition $P(B) > 0$

Il s'agit là du théorème de Bayes publié en 1763. Il apporte une relation de probabilités inverses : probabilité de la cause A , sur la base de l'effet observé B . $P(A_n)$ est la probabilité *a priori*, et $P(A_n|B)$ désigne la probabilité *a posteriori* de la cause A_n . Le mécanisme d'inférence bayésienne résulte de la combinaison de cette relation simple de probabilité avec la notion de temps qui amène le principe de causalité temporelle (cf Annexe sur les inférences bayésiennes).

Définition B.0.6. *Si chaque point échantillon ω représente la séquence $\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\}$, notée aussi $\{X_i, 1 \leq i \leq n\}$ on appelle une telle séquence un processus stochastique, i.e. une famille de variables aléatoires.*

Moyenne et variance

Une variable aléatoire X est *sommable* quand la condition (B.0.21) est satisfaite

$$E(|X|) < \infty \quad \text{avec} \quad E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) \quad (\text{B.0.21})$$

Dans ce cas X a une espérance ou moyenne finie.

Si X est bornée par M ,

$$|X(\omega)| \leq M \quad \forall \quad \omega \in \Omega$$

alors X est *sommable* et

$$E(|X|) = \sum_{\omega} |X(\omega)| P(\omega) \leq M \sum_{\omega} P(\omega) = M$$

En particulier, si Ω est fini, alors toute variable aléatoire définie sur Ω est bornée et

$$X \mapsto E(X) \tag{B.0.22}$$

assigne un nombre à la variable aléatoire.

Si A est un événement, son indicateur I_A est une variable aléatoire et

$$E(I_A) = P(A)$$

ainsi, la notion d'espérance mathématique étend celle de mesure de probabilité.

Théorème B.0.6. *Si X et Y sont sommables, alors il en est de même pour $X+Y$ et*

$$E(X + Y) = E(X) + E(Y) \tag{B.0.23}$$

De (B.0.23) il vient

$$E(a) = a \quad , \quad E(aX) = aE(X) \tag{B.0.24}$$

$$E(aX + bY) = aE(X) + bE(Y) \tag{B.0.25}$$

ce qui fait de (B.0.22) un opérateur linéaire.

Par extension de (B.0.6), si $X_1, X_2, \dots, X_n$ sont des variables aléatoires sommables alors

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) \tag{B.0.26}$$

Annexe C

Raisonnements probabilistes

C.0.2 Raisonnements probabilistiques

Nous présentons ici quelques principes théoriques et pratiques qui permettent d'élaborer et d'utiliser les modèles bayesiens. Nos sources sont principalement les ouvrages suivants : [Pearl \[1988\]](#) , [Jensen \[1996\]](#) , et [Jensen \[2002\]](#).

Approche extensionnelle et intentionnelle

Nous distinguons principalement deux approches pour les raisonnements en Intelligence artificielle : une **approche extensionnelle** et une **approche intensionnelle**. L'approche extensionnelle utilise un ensemble de règles dont les combinaisons et évaluation traitent l'incertitude comme une valeur de vérité généralisée : la certitude d'une *formule* est définie comme une unique fonction de la certitude de ses *formules* constitutives. Par contre, dans un système intensionnel, les mesures de certitude sont affectées à des ensembles issus d'univers différents, que l'on combine avec les opérations de la théorie des ensemble.

Le réseau de règles construit dans un système extensionnel est un ensemble de procédures aboutissant à la certitude d'une conclusion , étant donné la crédibilité des règles et de la certitudes des prémisses. La combinaison de *la crédibilité locale* (i.e. sans considération des autres éléments environnants) et du *détachement de l'origine* (i.e. sans considérer d'où viennent les faits) constitue le principe du concept de la *modularité procédurale*

du système extensionnel. Au contraire, dans un formalisme intensionnel, comme celui de la théorie des probabilités, un score probabiliste local nous laisse incapable de prendre une décision avant d'avoir vérifié l'irrélevance de tous les autres éléments de la base de connaissance. La vérification de l'irrélevance est un point crucial des systèmes intensionnels.

Les systèmes extensionnels présentent souvent des incohérences et des conclusions contre-intuitives : manipulations impropres des inférences bi-directionnelles (dans le sens direct et le sens inverse), difficultés de rétracter des conclusions, traitements impropres des éléments de preuve corrélés. Rappelons que l'inférence est l'opération qui consiste à admettre une proposition en raison de son lien avec une proposition préalable tenue pour vraie.

La possibilité d'utiliser à la fois une information de diagnostic et de prédiction est un point primordial des systèmes intensionnels. Ainsi le pattern courant de l' *induction* : si A implique B , alors trouver B vrai rend A plus crédible. Un système extensionnel ne permet qu'un raisonnement diagnostique, sans inférences prédictives. Cette absence de composantes prédictives retire au système un autre aspect important du raisonnement plausible : celui de ***l'explication a posteriori*** : si A implique B , C implique B , et B est vrai, alors trouver que C est vrai rend A moins crédible. Pour cela, le système doit utiliser des inférences bi-directionnelles : de *l'élément de preuve* vers *l'hypothèse* et de *l'hypothèse* vers *l'élément de preuve*, ou de l' *l'effet* vers la *cause* et de la *cause* vers l' *effet* : l' *effet* étant ce qui est produit par la *cause* qui peut être cachée, il est *l'élément de preuve* observé produit par *l'hypothèse* que l'on peut ignorer : en fait, l' *effet* est aussi *les faits* visibles de la *cause* invisible. On peut voir aussi, dans un lien avec les probabilités, *l'élément de preuve* comme un *effet incomplet*, et *l'hypothèse* comme une *cause partielle*. L'esprit de modularité procédurale de la logique déductive est alors révoqué dans un système intensionnel : pour le *détachement de l'origine* comme pour la *crédibilité locale* on arrive à des contradictions. Ainsi, indépendamment de l'échelle de mesure, numérique, binaire ou établie à partir d'autres qualificatifs, il semble y avoir à la base une lutte entre la modularité procédurale et la cohérence sémantique, peu importe la notation utilisée.

Un système extensionnel répond à l'intensité des poids et non à leurs origines, ils produisent les mêmes conclusions si les poids sont identiques et les origines diverses.

Dans un système intensionnel, la syntaxe consiste en des phrases déclaratives sur des

états de faits, ce qui rend compte assez bien de la connaissance de l'univers considéré. Les systèmes intensionnels utilisent sans problèmes les inférences bi-directionnelles et éléments de preuve corrélés. Néanmoins, comme la syntaxe ne met pas en évidence une procédure utilitaire de calcul, il faut construire des mécanismes spéciaux qui convertissent les entrées déclaratives en routines qui répondent aux questions : les *réseaux de croyance* (*belief network*) constituent ces mécanismes.

Le but est de réaliser un système intensionnel opérationnel en rendant les relations de dépendance (*relevance*) explicites, et de trouver un remède à l'impuissance d'une déclaration telle que $P(A|B) = p$. On ne peut agir avec cette seule déclaration à moins d'être sûr que les autres éléments de la base de connaissance n'ont rien à voir avec B et peuvent ainsi être ignorés. L'astuce est d'encoder la connaissance de manière à reconnaître ce qu'on ignore. Les réseaux de croyance représentent graphiquement les dépendances comme des noeuds voisins, permettant d'agir et de prendre une décision après consultation du voisinage. Peu importe ce qu'on ne peut voir localement, la représentation en réseau offre la mise à jour dynamique de tous les éléments qu'on a le droit d'ignorer, et délivre ainsi les permissions d'agir.

Les réseaux de croyance bayésien ont une sémantique claire : ils sont eux même une partie intégrale de la sémantique de la base de connaissance qu'ils modélisent. La théorie des probabilités est le formalisme qui joue un rôle central pour les réseaux de croyance bayésiens.

Le point important des systèmes intensionnels est comment articuler les conditions sous lesquelles un événement a à *voir* avec un autre événement. Différents formalismes existent, mais concernant les réseaux bayésiens, de par la théorie des probabilités, cette notion de "*avoir à voir avec*" est associée à la notion de **dépendance** dans son sens probabilistique. L'exigence essentielle nécessaire à la stabilité et à l'entièreté du réseau est sa **structure causale**, i.e. on doit identifier les antécédants les plus pertinents (ou dépendants) de chaque variable de manière récursive, dans un ordre temporel total.

L'utilisation des probabilités a pour but de donner un résultat quantitatif cohérent sur l'évolution de la croyance d'un événement à la lumière d'une information partielle ou incertaine. Néanmoins, la probabilité n'est pas tant au sujet des nombres mais plutôt de la structure du raisonnement.

La correspondance heureuse entre l'intuition humaine et les lois de la proportion n'est

pas une pure coïncidence. **Les croyances ne se forment pas dans le vide, elles proviennent des expériences sensorielles.** Nous oublions souvent les expériences et gardons les impressions mentales sous forme de moyennes, de poids, ou encore plus vivement de relations qualitatives abstraites qui nous aident à déterminer des actions futures. De même, nos calculs de croyance évoluent à partir des calculs de proportions et de fréquence d'occurrences, i.e. des probabilités.

Nous distinguons quatre relations primaires dans le langage des probabilités :

- **vraisemblance** (*likelihood*) ("il est plus *vraisemblable* que Tim vole plutôt qu'il marche", ou "Tim est *plus à même* de voler que de marcher")
- **condition** ("si Tim est malade, il ne peut pas voler")
- **relevance** ("le fait que Tim vole *relève de* son état de santé")
- **causalité** ou **cause** ("être malade est la *cause* de l'inaptitude de Tim à voler")

Condition

L'affirmation : "étant donné que ce que je connais est C " ou "sachant C " se note dans le formalisme des probabilités de la façon suivante : $P(A|C) = p$. Cette affirmation combine à la fois les notions de connaissance et de croyance en attribuant à A un degré de croyance p , étant donné la connaissance de C . C est aussi appelé le contexte de connaissance de A . Thomas Bayes (1702-1761) fut le premier à associer cette affirmation avec la fraction suivante devenue la définition de la notion de probabilité conditionnelle :

$$P(A|C) = \frac{P(A, C)}{P(C)} \quad (\text{C.0.1})$$

La **dynamique de mise à jour de la croyance** sous des contextes changeants n'est pas arbitraire mais doit obéir à une loi de plausibilité basique, comme celle du *milieu hypothétique* ainsi défini :

$$P(B|C) = \alpha P(B|C, A) + (1 - \alpha) P(B|C, \neg A) \quad (\text{C.0.2})$$

avec $\alpha = P(A|C)$ est un nombre entre 0 et 1

En effet , en posant $C = CA \cup C\neg A$, on a

$$\begin{aligned}
 P(B|C) &= \frac{P(BC)}{P(C)} = \frac{P(B.(CA \cup C\neg A))}{P(C)} \\
 &= \frac{P(BCA \cup BC\neg A)}{P(C)} = \frac{P(BCA)}{P(C)} + \frac{P(BC\neg A)}{P(C)} \\
 &\text{car } BCA \text{ et } BC\neg A \text{ sont des ensembles disjoints} \\
 P(B|C) &= \frac{P(BCA)}{P(CA)} \frac{P(CA)}{P(C)} + \frac{P(BC\neg A)}{P(C\neg A)} \frac{P(C\neg A)}{P(C)} \\
 &= P(B|CA)P(A|C) + P(B|C\neg A)P(\neg A|C) \\
 &= P(A|C)P(B|CA) + (1 - P(A|C))P(B|C\neg A) \\
 &= \alpha P(B|C, A) + (1 - \alpha)P(B|C, \neg A)
 \end{aligned}$$

□

Relevance

La relevance indique un changement possible de la croyance d'un événement, dû à un changement spécifique survenu dans le contexte de connaissance. Deux propositions A et B relèvent l'une de l'autre dans un contexte C si le fait d'ajouter B à C modifie la vraisemblance de A . On peut définir la relevance en terme de vraisemblance et de condition, mais c'est une notion plus basique que la vraisemblance : s'il est difficile d'estimer la vraisemblance de 2 événements, il est souvent plus facile de juger leur relevance réciproque.

Causalité

La causalité conduit au schéma de dépendance suivant : deux événements ne deviennent pas pertinents l'un de l'autre par le seul fait de leur prédire une conséquence commune, mais ils le deviennent quand la conséquence est actuellement observée.

Inférence bayésienne

Dans le formalisme bayésien, la croyance suit les trois axiomes élémentaires de la théorie de la probabilité :

$$\begin{aligned}
 0 &\leq P(A) \leq 1 \\
 P(\text{proposition sûre}) &= 1 \\
 P(A \text{ ou } B) &= P(A) + P(B) \text{ si } A \text{ et } B \text{ sont mutuellement exclusifs, i.e. } A \cap B = \emptyset
 \end{aligned} \tag{C.0.3}$$

Comme on peut toujours écrire un événement A sous la forme de l'union des événements joints $(A \cap B)$ et $(A \cap \neg B)$, on a alors

$$P(A) = P(A, B) + P(A, \neg B) \quad (\text{C.0.4})$$

Plus généralement, si $B_i, i = 1, 2, \dots, n$, est un ensemble exhaustif de propositions exclusives (i.e. une partition de B), alors

$$P(A) = \sum_i P(A, B_i) \quad (\text{C.0.5})$$

Une conséquence directe de (C.0.3) et (C.0.4) est d'assigner la valeur unité à la croyance totale d'une proposition et de sa négation,

$$P(A) + P(\neg A) = 1 \quad (\text{C.0.6})$$

$P(A|B)$ spécifie la croyance en A sous l'hypothèse que B est connue de façon certaine, on note

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{P(AB)}{P(B)} \quad (\text{C.0.7})$$

Si $P(A|B) = P(A)$, on dit que A et B sont indépendants.

Si

$$P(A|B, C) = P(A|C) \quad (\text{C.0.8})$$

on dit que A et B sont conditionnellement indépendants sachant C , ou bien indépendants l'un de l'autre dans le contexte C :

$$\begin{aligned} P(A|B, C) = P(A|C) &\Leftrightarrow \frac{P(ABC)}{P(BC)} = \frac{P(AC)}{P(C)} \\ &\Leftrightarrow \frac{P(ABC)}{P(AC)} = \frac{P(BC)}{P(C)} \\ &\Leftrightarrow P(B|A, C) = P(B|C) \end{aligned}$$

□

Dans l'expression :

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad (\text{C.0.9})$$

B est le contexte ou cadre de connaissances, et $A|B$ est l'événement A dans le contexte B . En conséquence, les connaissances factuelles sont encodées sous la forme de probabilités conditionnelles, et la croyance d'événements joints calculée à l'aide du produit :

$$P(A, B) = P(A|B)P(B) \quad (\text{C.0.10})$$

d'où il vient, avec (C.0.5) :

$$P(A) = \sum_i P(A|B_i)P(B_i) \quad (\text{C.0.11})$$

Si K désigne l'hypothèse commune du contexte de connaissance, nous avons :

$$P(A|K) = \sum_i P(A|B_i, K)P(B_i|K) \quad (\text{C.0.12})$$

en effet ,

$$A = \sum_i A \cap B_i \quad , \text{ et dans le contexte } K \text{ explicité : } A_{(K)} = \sum_i A_{(K)} \cap B_{i(K)}$$

$$\text{d'où } P(A|K) = \sum_i P(A|K, B_i|K) = \sum_i P(A|K \mid B_i|K)P(B_i|K) = P(A|B_i, K)P(B_i|K)$$

□

Il vient également, avec (C.0.10), pour n événements, $E_1, E_2, \dots, E_n$, que la probabilité de l'événement joint $(E_1, E_2, \dots, E_n)$ peut s'écrire comme un produit de n probabilités conditionnelles :

$$P(E_1, E_2, \dots, E_n) = P(E_n|E_{n-1}, \dots, E_2, E_1) \dots P(E_2|E_1)P(E_1) \quad (\text{C.0.13})$$

Dans le cas de plusieurs variables *conditionnellement indépendantes sachant H* , on peut montrer avec (C.0.8) et (C.0.13) que

$$P(E_1, E_2, \dots, E_n|H) = \prod_{k=1}^N P(E_k|H) \quad (\text{C.0.14})$$

Le coeur de la technique bayésienne vient de la célèbre formule d'inversion :

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)} \quad (\text{C.0.15})$$

qui énonce que la croyance accordée à l'hypothèse H d'après l'observation e peut être calculée en multipliant la croyance précédante $P(H)$ par la vraisemblance $P(e|H)$ que e se réalise

si H est vrai. On appelle $P(H|e)$ la *probabilité a posteriori* et $P(H)$ la *probabilité a priori*. Le dénominateur $P(e)$ de (C.0.15) est considéré comme un facteur de normalisation $P(e) = P(e|H)P(H) + P(e|\neg H)P(\neg H)$ et peut être calculé en exigeant que $P(H|e) + P(\neg H|e) = 1$

On peut donc regarder (C.0.15) comme une règle normative pour mettre à jour la croyance en réponse à l'observation.

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad \text{et} \quad P(B|A) = \frac{P(A, B)}{P(A)} \quad (\text{C.0.16})$$

(C.0.16) signifie, entre autre, que la croyance qu'une personne attribue à B ***après la découverte de A*** , i.e. $P(B|A)$, n'est jamais inférieure à la croyance attribuée à (A et B) ***avant la découverte de A*** , i.e. $P(A, B)$. Le rapport de ces deux croyances augmente proportionnellement avec le degré de surprise $[P(A)]^{-1}$ associé à la découverte de A .

Habituellement, la méthode de calcul des modèles probabilistiques utilise une fonction de distribution jointe, qui assigne des poids positifs à chaque événement élémentaire de telle sorte que leur somme soit égale à 1 : si A, B, C sont les trois événements élémentaires, la fonction de distribution de probabilité jointe associera des poids à $P(ABC), P(AB\neg C), \dots, P(\neg A\neg B\neg C)$ de telle sorte que leur somme soit égale à 1. Néanmoins, en pratique, les fonctions de distribution jointe des probabilités sont rarement explicitées, dans l'analyse des variables aléatoires continues, les distributions ont une expression algébrique, et pour les variables aléatoires discrètes, des méthodes de représentation indirectes ont été développées, où la distribution générale est inférée à partir de relations locales parmi de petits groupes de variables : c'est ce que permet l'approche en réseau.

Diagnostic et prédiction

La nature de (C.0.15) peut s'interpréter avec les paramètres de *cote* et de *rapport de vraisemblance*. En divisant (C.0.15) par la forme complémentaire pour $P(\neg H|e)$:

$$\frac{P(H|e)}{P(\neg H|e)} = \frac{P(e|H)}{P(e|\neg H)} \frac{P(H)}{P(\neg H)} \quad (\text{C.0.17})$$

on définit la *cote a priori* sur H par

$$O(H) = \frac{P(H)}{P(\neg H)} = \frac{P(H)}{1 - P(\neg H)} \quad (\text{C.0.18})$$

le rapport de vraisemblance par

$$L(e|H) = \frac{P(e|H)}{P(e|\neg H)} \quad (\text{C.0.19})$$

et la *cote a posteriori*

$$O(H|e) = \frac{P(H|e)}{P(\neg H|e)} \quad (\text{C.0.20})$$

est donnée par le produit

$$O(H|e) = L(e|H)O(H) \quad (\text{C.0.21})$$

Finalement, la confiance globale de la croyance en l'hypothèse H , en connaissant le contexte K et l'observation e , apparaît comme le produit de deux facteurs : la *cote a priori* $O(H)$ et le *rapport de vraisemblance* $L(e|H)$. Le premier facteur indique de manière prédictive ou prospective la confiance accordée à H , alors que le second représente le diagnostic ou la confiance rétrospective accordée à H après observation de e .

Un des aspects intéressants du calcul bayésien est qu'il permet un schéma de calcul récursif et incrémental, si H dénote l'hypothèse et $e_n = e^1, e^2, \dots, e^n$ une séquence de données observées dans le passé, pour un fait nouveau e , on calcule pour la nouvelle séquence $e_{n+1} = \{e_n, e\}$,

$$P(H|e_n, e) = P(H|e_n) \frac{P(e|e_n, H)}{P(e|e_n)} \quad (\text{C.0.22})$$

En effet,

$$\begin{aligned} P(H|e_n, e) &= \frac{P(e, e_n, H)}{P(e_n, e)} \\ &= \frac{P(e|e_n, H)P(e_n, H)}{P(e|e_n)P(e_n)} \\ &= \frac{P(e|e_n, H)P(H|e_n)P(e_n)}{P(e|e_n)P(e_n)} \\ &= P(H|e_n) \frac{P(e|e_n, H)}{P(e|e_n)} \end{aligned}$$

□

En rapprochant (C.0.22) de (C.0.15) nous voyons que la croyance antécédante $P(H|e_n)$ devient maintenant la probabilité a priori (elle résume l'expérience passée). La mise à jour se fait alors en la multipliant par la fonction de vraisemblance $P(e|e_n, H)$ qui mesure la probabilité de la nouvelle donnée, connaissant l'hypothèse et les observations passées.

Sous la condition que la vraisemblance soit indépendante des données passées et qu'elle implique seulement e et H , on peut écrire

$$P(e|e_n, H) = P(e|H) \quad \text{et} \quad P(e|e_n, \neg H) = P(e|\neg H) \quad (\text{C.0.23})$$

On obtient en divisant (C.0.22) par l'équation complémentaire pour $\neg H$

$$O(H|e_{n+1}) = O(H|e_n)L(e|H) \quad (\text{C.0.24})$$

Ainsi quand une donnée nouvelle arrive, on multiplie la *cote a priori* courante $O(H|e_n)$ par le *rapport de vraisemblance* de e . La nature incrémentale de la mise à jour du processus apparaît sous la forme logarithmique :

$$\log O(H|e_{n+1}) = \log O(H|e_n) + \log L(e|H) \quad (\text{C.0.25})$$

Attention, la validation du processus de mise à jour récursive dépend fortement des conditions d'indépendance des informations *passées* par rapport aux informations *futures* : elles doivent leur être complètement irrelevantes.

Annexe D

Publication d'article

Red Light Spotters

Images-Driven Sound and Rhythm Landscape

Philippe Codognet
Keio University, DMC

Research Institute for Digital Media and Content
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
philippe@dmc.keio.ac.jp

Gilbert Nouno

University of Paris 6, LIP6
IRCAM

1 place Igor Stravinsky, 75004 Paris, France
Gilbert.Nouno@ircam.fr

ABSTRACT

Red Light Spotters is a multi-contextual interactive project which aims to create an open and real-time musical environment sharing relations with and extending a natural ambient landscape. It encompasses artistic creation process embedding image tracking, artificial intelligence with inductive original tempo tracking and beat prediction algorithms. We end up with emergent rhythmic processing for ambient musical creation, interpreted and performed by sound artists.

Categories and Subject Descriptors

H.5.5 [Sound and Music Computing]: ModelingSignal analysis, synthesis, and processing; H.5.1 [Multimedia Information Systems]: Audio input/output

General Terms

Design, Algorithms, Experimentation

Keywords

Interactive sound installation, sensor data interpretation, digital art, computer-based music, creation techniques

1. INTRODUCTION

Red Light Spotters is a multi-contextual interactive project which aims to create an open and real-time musical environment sharing relations with and extending a natural ambient environment, namely an urban landscape. This interactive sound installation is intended to be exhibited in a "City View" observatory that are becoming more and more frequent in top floors of skyscrapers buildings, and to be an auditory complement to the visual stimulus from the elevated point of view over the city. In a first step, this project is specifically developed for Tokyo, as there the blinking red lights on top of buildings are creating a mesmerizing dynamic virtual landscape and can be used to create rhythms and music. But it can be further adapted

for other cities. It encompasses artistic creation process embedding image tracking, artificial intelligence with inductive original tempo tracking and beat prediction algorithms. We perform pertinent image and symbolic descriptors extractions, as pulsation and rhythm features, in order to synchronize both musical and control worlds with the natural visual environment. We end up with emergent rhythmic process for musical creation interpreted and performed to a certain extent by sound artists. This project is based on an idea by multimedia artist Maurice Benayoun and researcher Philippe Codognet and is developed by a collective of sound artists and researchers¹ as part of **Keio University DMC**²'s artistic and research projects.

In this paper we will present in Section 2 the natural visual framework of our project. We will then move to the constitutive elements and flowing data description of the installation in Section 3. Some details about the algorithmic intelligence linking the input data to the musical generation are discussed in Section 4. We will argue about the interpretation and interaction aspects in Section 7. Finally, the implementation of the computer environment is presented in Section 5 and the musical composition environment is considered in Section 6. A brief conclusion ends the paper.

2. THE VISUAL POINT OF VIEW

The town of Tokyo at night Fig.1(a), with its specific blinking red lights skyline ambiance³, is the first experimental urban setup that we propose to use as an immersive visual and musical experience. Our concept is aimed to create intelligent real-time relationships between the context of the installation and the exhibition space itself; that is, between the environmental images of the town (i.e., somehow a landscape of lights) and the interactive musical surrounding occurrences and resonances. The spectators will thus be immersed in these related images and sounds and intuitively link them together, as the music is created by the city itself. However, we do not want to make this relationship too much explicit for the spectators, as this is not an "interactive technology demo" but a sound artwork, therefore our goal is to have the spectators first appreciating the musical ambience as such and then to have them slowly realizing the relationships and interactions between the city lights and the sound rhythms. We also wish to leave the music landscape and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAME'08, October 31, 2008, Vancouver, British Columbia, Canada.

Copyright 2008 ACM 978-1-60558-314-3/08/10 ...\$5.00.

¹Alice Daquet, Gilbert Nouno, Philippe Codognet

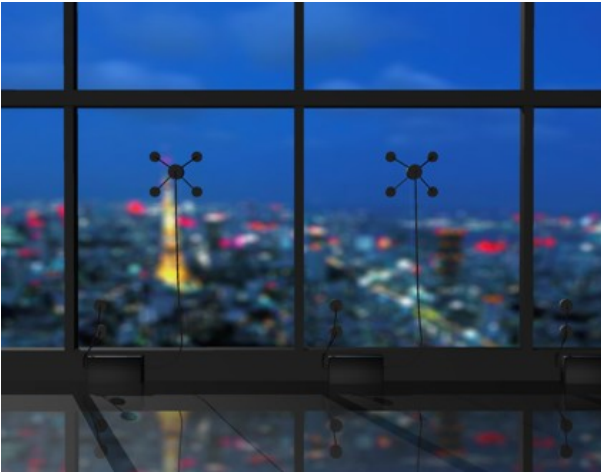
²research institute for digital media and content

³those blinking red lights are used to signal the top and sides of various building and skyscrapers to helicopters flying over the city

composition tool open in order to renew itself in different contexts depending on the location, the time and the sound artists.



(a) The actual landscape as could be seen from the "Tokyo City View" observatory on the top of Mori Tower building, 52nd floor



(b) Tokyo view with detail of camera installation for image acquisition

Figure 1: Tokyo city view

3. FLOWING IMAGES TOWARDS SOUND

In this section we describe the real-time processing of the data flux and the transformations they undergo in the installation space. The emphasis is put on the perceptual experience that should result from this process, as the visual landscape would be the initial element offered to the eyes of the viewers who become listeners and potentially actors of their own multi-sensorial experience. By this, we aim to extract or reveal a certain amount of information handled in the image - the blinking red lights - and turn this information into musical output that we can control and modulate artistically.

The initial perceptive data is the dynamic view of the city with its intricate mesh of flashing red lights. It is turned by video camera grabbing into a series of image frames that we

reduce after image-signal processing into a set of blinking dots (Fig.2) with characteristic parameters such as intensity, frequency and time-phase. These parameters extractions - features or descriptors - lead us to some rhythmic considerations and relations that we can achieve through different ways of grouping and correlating the blinking lights. We end up with an interdependent rhythmic structure in relation with the expectations that the viewer might have while contemplating the natural view offered to his eyes.

We relate the expectation of the spectator with some elements of anticipation: we use in an inner-predictive way parameters in the computation of the predictive algorithm that will be presented later in section 4. We implement the algorithmic process on the temporal flowing sequence of the flashing red lights so as to produce musical events synchronized with the quasi-random visual occurrences. Doing so, the musical content consists of real-time generated rhythms and sounds based on some predictability related to the quasi-random visual dots components. By adding very simple rules to the output of this on-the-fly-created compositional environment, we bring on emergent musical process [13]. We call the red lights quasi-random elements as the lights might have a quasi-constant time blinking interval - also called inter-onset if speaking about musical data. The way we decide to group the observed flashing dots can be dynamically modified and has a major influence on the tempo/synchronization decoding part.

The tempo/synchronization decoding process is based on an inference stochastic framework which translates the visual impulses and turns them into musical rhythmic and textural elements. The way to look at the skyline landscape is manifold because of the too numerous lights, and thus we also keep a varied and open-ended interpretative musical environment that can be turned into either a compositional, improvisational or interpretative context depending on the choices made at a given time by the composer/interpreter of the musical session. Our musical environment is open to different sound artists and musical styles.

We are also considering in a further extension about the possibility of inviting the public to be part of the interpretative musical process. The quality of the interaction we are looking for is not only a simple and basic human-machine interaction, but rather a human to human interaction through machine technology. This idea is important as it considers the technology as a medium delivering and being the information as the same time. The point of view on the machines and their use evolves to a doubtless interactive context involving spectators at the foreground. In this installation, the technology is a factor of communication but does not receive the main focus, as opposed to so many digital art installations which focus on technology more than on the artistic process itself. We situate this project in a new media and technology-dependent way of conceiving and producing artistic processes and also relations between spectators, as exemplified in many work of contemporary art in the last decades [3]. This underlies that the sensorial emotion resulting from the attending experience is more important than the technology itself. Because computer technology and electronic-devices have now become basic elements of everyday life in the emerging ubiquitous society, we aim at designing with them and by them a contextual ambient media experience which might underline and emphasis the natural and interactive links between vision and audition.

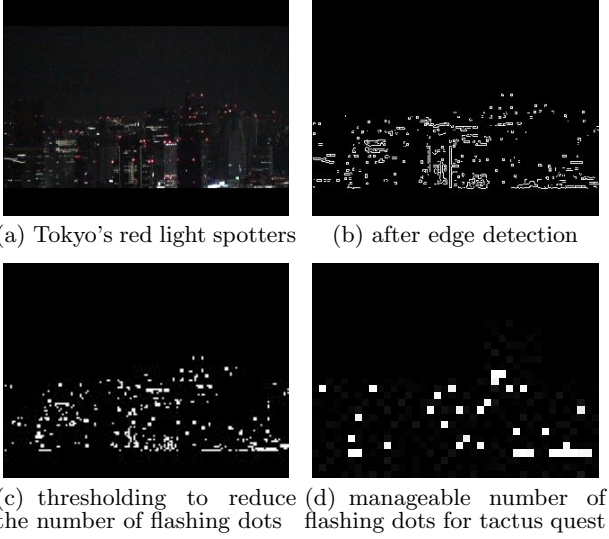


Figure 2: Image processing on Tokyo video capture

4. FLASHING DOTS FOR TEMPO

We will discuss now about some more theoretical aspects that turn the red light spotters into a temporal flow of pulsations needed to drive a musical tempo. We recall some concept of musical time and come up with the foundations of an algorithm for tempo tracking and pulsation prediction regarding separate dots or groups of flashing dots.

4.1 Musical time

The notion of musical time, as simple as it seems to be, can be a surprisingly difficult concept to grasp. Music is composed of events which unfold within flowing time. The way they are organized creates a perceptual sensation of musical time. This organized time oscillates between two extremes, one completely loose and the other completely straight in its organization. Some composers refers to this quality of time as a continuous time and a striated time [2], and even if applied to harmony in a general sense of the music, it applies very well to the musical time. The reality of musical time lies between these two extreme boundaries. What we are interested in is the distribution of time which is closer to the striated time - namely a pulsed time - in which we can perceive a more or less strong cyclic sensation of organized time that we interpret as a pulsation and then as a tempo.

4.2 Input data and cognitive hypothesis

We present in the following the theoretical framework, with input and output data, and the method based on an optimal grid search we are proposing to solve the problem of pulsation estimation and tracking.

We consider the domain of symbolic events occurring in a series of onset time $\{t_n\}$ and focus our research on a temporal set of events which represent the instant polyphony or poly-rhythms of flashing dots. We can have the same approach if we consider the events as onset extracted from audio signal [4] [8].

The main problem to solve is to estimate in real time a pulsation which can be considered as the best matching pulsation of a temporal input sequence. This also imply there is not a single and unique master pulsation representative of the temporal data. Considering phrases of rhythm, this aspect is confirmed by the operation of transcription, when asking people to notate a rhythm they are listening to. When performing the imitation of musical tempo, it is often common to experience octave errors. Simha Arom [1] names one of these best underlying pulses the *tactus*, he also presents the incredible amount of confusing names trying to describe different and not always clear notions related to tempo. The *tactus* is presented as the greatest of the smallest pulses that *inducts* the *tempo*, and which enable the rhythmic structure to unfold on a sometimes unheard and hidden temporal grid. The *tactus* represents an underlying subdivided pulsation you can beat and which enables you to built and understand all the rhythms that are played. This consideration is one of the hypothesis of our research : if there is perception of a tempo, we presume there is a regularity which could be revealed by a filigree grid made by the regular *tactus* pulses. Of course this *regularity* is more or less accurate. When performing music, the tempo is moving, either because of unwanted inaccuracy, or because of a specific musical style. When extracting pulses from flashing dots considered as random variables, it is obvious that the resulting tempo, if found, would likely be non constant and depending on the (possibly changing) group of observed dots. The train of random incoming red-light-pulses is analyzed in real time. The observations are converted into some likelihood's of probable tempi. A Hidden Markov Model (HMM) framework is then use to decode the tempo estimation. The result is a tracking of the tempo and a prediction of the pulsation which can be used as a variable tempo-track to drive some musical hardwares or softwares modules.

4.3 Tactus and quantification

Considering the above hypothesis of perceptual regularity inducting a tempo, we are looking for a regular grid that can materialize the tempo. If we make this underneath grid appear, we also propose a quantification of the temporal events. This quantification is one of the solutions of the rhythmic transcription of the input phrase.

To consider onsets of the input sequence, we threshold the intensity of the flashing dots coming out of the image matrix analysis Fig.2(d): when going above this threshold in a cyclic time, the trigger generate successive attacks, from which inter-onsets.

4.4 Paradigm for the tactus search

[14] proposes a tactus decoding approach based on the inter-onset set $\{o_i\}$ of a time events sequence. In an ideal and mechanical input data case, the tempo is absolutely still, and the *tactus* would result as a *greater common divisor* (*gcd*) of the inter-onset. For a real case, [14] considers the inter-onset to have random deviations, and defines an error function of the *tactus* period Δ and $\{o_i\}$:

$$e(\Delta) = \sum_{i=1}^n \left(\frac{o_i}{\Delta} - \left\lfloor \frac{o_i}{\Delta} + \frac{1}{2} \right\rfloor \right)^2 \quad (1)$$

local minimum of Eq.(1) are potential candidates for the

tactus. If an exact gcd exists, it is given by

$$gcd(o_1, o_2, \dots, o_n) = \max\{\Delta | e(\Delta) = 0\}$$

The best candidates of Eq.(1) remains approximated as an empiric threshold is used to select the local minimum. Moreover, the graph of Eq.(1) can be unclear and does not always provide a best estimated candidate, namely Δ^* . The reason for this is that we can't differentiate Eq.(1) because of the integer part. We propose a slightly modified equation to replace Eq.(1) and keep the idea of *tactus* search which is widely used in music transcription [7].

4.5 Optimal grid search

This section regroups the mathematical concepts to solve the quest of the *tactus*. The original *Optimal Grid* approach discussed here is inspired by operational research mathematician Pr. Bruno Simeone [15]. We have chosen not to detail in this paper the proofs of some more technical mathematical results due to the limited space.

4.5.1 Mono-dimensional problem

Let's take a time onset series $\{t_1, \dots, t_n\}$. Suppose a regular best approximation grid of step Δ and of length $p + 1$ points which includes the origin : $0, \Delta, 2\Delta, \dots, (p-1)\Delta, p\Delta$. We can show that we are looking for Δ^* such that it minimizes the following equation

$$f(\Delta) = \sum_{k=1}^n \min_{h=0,1,\dots,p} |t_k - h\Delta| \quad (2)$$

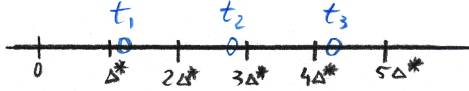


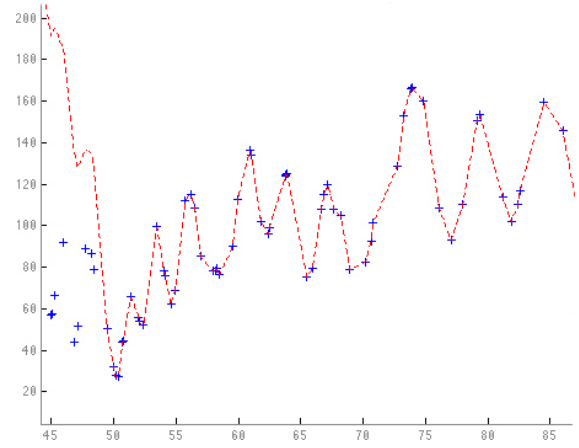
Figure 3: grid example with $k = 3$ et $p = 5$

We demonstrate that in a given close definition domain, Eq.(2) is the sum of n breakpoint functions. We demonstrate also that Eq.(2) admits $n(2p-1)$ singular breakpoints we note α_{kh} . As f is a breakpoint linear function, it finds its minimum among the α_{kh} . Through this result, we built an $\mathcal{O}(n^2(2p-1))$ procedure which gives all the local minimum and maximum of f . We are now sure that Δ^* is among the finite number of candidates which consist of the local maxima and minima. We have solved Eq.(2) without thresholding. p , the length of the grid, is also a parameter that can be seen as a *quantification resolution* for the transcription. Figure 4(a) show a reconstruction of f with singular breakpoints in comparison with its numerically approximated computation.

4.5.2 Phase of grid and bi-dimensional problem

The Δ step grids we have found above include the origin point. But there is no requirement for the optimal grid to include the origin point. This leads us to extend Eq.(2) to a bi-dimensional equation of variables *step* Δ and *phase* or *center* s .

Let's consider a unidimensional grid of step Δ and center $s : \{\dots, s-2\Delta, s-\Delta, s, s+\Delta, s+2\Delta, \dots\}$. For a given Δ , we want to find a center s^* which minimizes the grid phase function as the following :



(a) building f with $p = 26$. the best Δ^* candidate is 50.4. The breakpoints are the +, dashed-line is the numerically approximated function

Figure 4: singular breakpoints of an onset $\{100, 203, 406, 702, 989, 1310\}$ sequence

$$\varphi(s) = \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - h\Delta| \quad (3)$$

φ is Δ -periodic. We show that it reaches its minimum in a point of the form $s^* = t_q - l\Delta$ with $q \in \{1, \dots, n\}$ and $l \in \{m_q, \dots, M_q\}$ related boundaries we can also define.

We prove then the following theorem:

THEOREM 1. *Let's consider a Δ -step grid with center s^* which best coincides with a series of n points $\{t_1, \dots, t_n\}$. Then there is at least one integer q^* , $1 \leq q^* \leq n$, such that t_{q^*} belongs to the grid.*

We will now use *optimal coincidence* or *best approximation* as equivalent terms.

From what precedes, we can reduce to a surface I the space area where to look for a minimum of the bi-dimensional function :

$$\varphi(s, \Delta) = \sum_{k=1}^n \min_{h \in \mathbb{Z}} |t_k - s - h\Delta| \quad , \quad \Delta \in I \quad (4)$$

where I is a truncated triangle surface. Figure 5(a) and 5(b) show the periodicity observation and an overview of the local minimum and maximum of the *indices of coincidence* - indicators of how well a grid quantifies the input points - of Eq.(4) in its reduced definition domain.

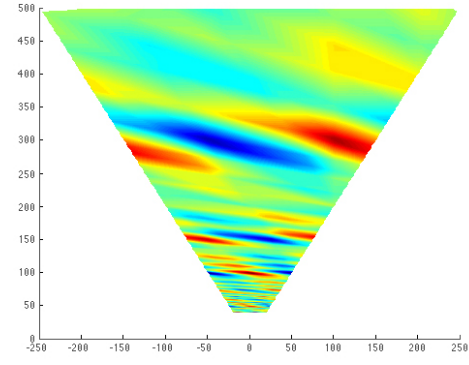
4.5.3 Bi-dimensional coincidence problem

With theorem Th.(1), we propose an $\mathcal{O}(n^3 p^2)$ procedure to solve Eq.(4). We give in the Alg.(1) pseudo-code the resulting bi-dimensional optimization algorithm.

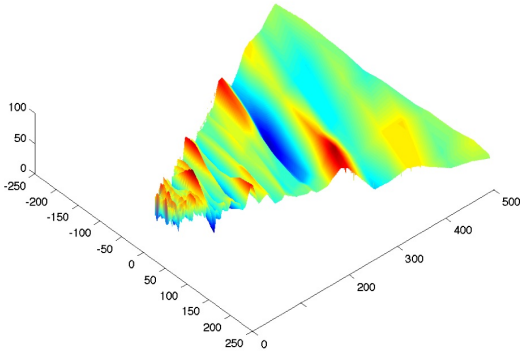
4.6 Probabilistic model

4.6.1 Observations and optimal grid search

The optimal grid search procedure provides us with some observations, namely some good candidates $\{\Delta_i\}$ and associated phase $\{s_i\}$. In dynamic programming environment,



(a) periodicity of φ on $\mathcal{I}(s, \Delta)$ and indices of coincidence



(b) detail for the optimal couple (s, Δ) with indices of coincidence

Figure 5: Approximation of φ on $\mathcal{I}(s, \Delta)$ for the sequence $\{100, 203, 406, 702, 989, 1315\}$

regarding a given observation window length, we obtain a series of observation vectors $\vec{\Delta}$ we consider as local density probability function.

4.6.2 Probabilistic HMM model

In a real-time context, we are interested in estimating and following the tempo to predict the next pulse. We use a Hidden Markov Model (HMM) to decode and infer the tempo-states transitions. The optimal grid paradigm provides us with some dynamic states sequences, updated on each new onset arrival. The unknown state is the tempo, given by Δ . We select some Δ candidates under the *index of coincidence* criterion. They provide the *likelihood probability* of the associated tempo for the observed sequence of time onset. This selection is done to avoid too many states among which too many would be improbable. It reduces also the amount of computations we need for the real-time implementation. With the selected Δ candidates we build on-the-fly a transition matrix between the states. We fill in the states transition probability matrix with gaussian probabilities centered on each candidate with respect to the others. This means the distance between two states affects directly their transition probability. We also use a $\log_2$ scale for the tempo in order to keep equal probabilities for acceleration

Algorithm 1 Bi-dimensional best approximation grid

Require: $\{t_1, \dots, t_n\}$ sequence of n input points

p number of points of grid plus center point

```

1: for  $q = 1$  to  $n$  do
2:   for  $l = 0$  to  $p$  do
3:     for  $k = 1$  to  $n$  do
4:       for  $h = 0$  to  $p, h \neq l$  do
5:          $\Delta_{kh} \leftarrow \frac{t_k - t_q}{h - l}$ 
6:          $\Phi_{kh} = \sum_{j=1}^n \min \left\{ \text{frac}(t_j - t_q, \Delta_{kh}), \Delta_{kh} - \right.$ 
            $\left. \text{frac}(t_j - t_q, \Delta_{kh}) \right\}$ 
7:       end for
8:     end for
9:      $\Phi_{ab} \leftarrow \min_{h,k} \Phi_{hk}$ 
10:     $\Delta \leftarrow \Delta_{ab}$ 
11:     $\Phi_{ql} \leftarrow \sum_{k=1}^n \min_{r=-l, \dots, p-l} |t_k - (t_q + r\Delta)|$ 
12:  end for
13: end for
14:  $\Phi_{q^*l^*} \leftarrow \min_{q,l} \Phi_{ql}$ 
15:  $\Delta^* \leftarrow \Delta_{q^*l^*}$ 
16:  $s^* \leftarrow t_{q^*} - l^* \Delta^*$ 
17: return  $s^*, \Delta^*$ 

```

and deceleration. The variance is chosen so as to allow more or less narrow tempo changes, a typical value being around 1% of the current tempo.

4.6.3 Tempo decoding, inference and prediction

Once we have defined the probabilistic model, we decode it with a forward Viterbi algorithm [12]. If we have no information on the initial tempo - as it would be mostly probable with the random flashing dots - we assume an initial flat prior. Running the HMM decoding in a dynamic programming environment gives us the tempo estimation and in the same time the phase of the *tactus* pulsation. As a direct consequence we can predict the anticipated *tactus* pulsation. It is interesting to note that the intrinsic phase computation of the optimal grid algorithm is used straight-forward as soon as we have made a decision for the tempo-state. To turn our paradigm into an anticipative behavior mode, we use the predicted pulse onset in the computation of the next incoming data : incorporating a future and forecast onset event into the present computation bring the algorithm in the anticipation domain [6].

4.6.4 Tempo tracking real-time implementation

We have chosen the real-time *MaxMSP*⁴ and *PureData*⁵ audio/control environment to implement the optimal grid algorithm as an object named *antepulse*, referring to pulse anticipation. The object is written with the *flex* C++ facilities programming library which enables to share the code for *OS X* and *Windows* computers, both for *MaxMSP* (Fig.6) and *PureData* environments.

⁴<http://www.cycling74.com/>

⁵<http://puredata.info/>

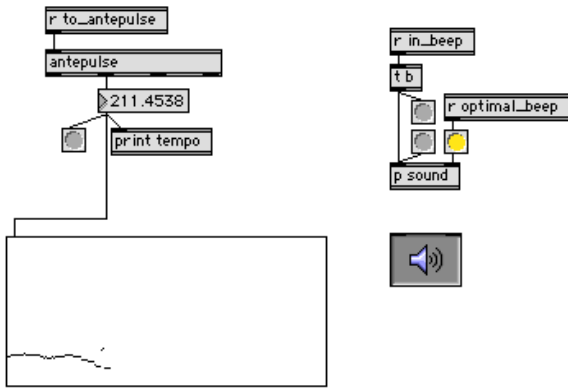


Figure 6: Antepulse in the *MaxMSP* environment

5. REAL-TIME FRAMEWORK

We present in this section some of the technical choices made in the *Red Light Spotters* installation.

The image analysis and its rhythmic decoding is done in the real-time *MaxMSP* environment. The image processing makes use of a specific edge detection algorithm [16] implemented in the *MaxMSP* graphic library *Jitter* in [10]. We end up with some rhythmic flux that can be used to drive other musical environments such as *Ableton Live*⁶, *Reaktor*⁷, *MaxMSP* or others, provided they support the *OSC*⁸ communication protocol, which enable dataflow through Ethernet protocol. The temporal decoded pulsation is sent by *MaxMSP* as a tempo-track for tempo synchronization, or can be received by *OSC* Plug-Ins in some clients applications, so as to be used freely for tempo and music generation. As the light might vary through the night, and as the phases of the lights are always moving, the resulting tempo is also moving accordingly to the input pulses, creating new variations of tempo and rhythms.

6. MUSIC AND STYLES

The musical hybrid compositional and interpretative system we propose is open in the sense that it is independent of any given music genre or style. Our intent is to let the musician/artist free to use the sounds, samples, audio effects and musical textures and combine them as he likes. We are providing a framework where this musical process is easy to work with. Concerning the performative aspects, several control devices such as touch screens could be used to create an interactive interpretation of the mixing or sound control processes.

By the use of basic interaction rules applied to the parallel decoded tempo - like reaching some point together, being close to the same speed etc - we achieve some emergent behaviors which generate some musical forms.

It has to be noted that due the rhythmic sound-generation process, the most obvious music that can be created with *Red Light Spotters* is clearly in the spirit of Steve Reich' minimal music, e.g. early works *Piano phase* (1967) or later compositions such as *Six marimbas* (1973-1986) and

Music for 18 musicians (1974-1976), although we are currently investigating a more contemporary style, also influenced by electro and dub style. Other musical styles could of course be investigated, but will require more development time in order to adapt and connect the rhythm generators to meaningful musical processes. Also it is worth noticing that as the rhythm generators used in *Red Light Spotters* can be considered as quasi-random, another influential style is the so-called random music, as pioneered in the 50's by John Cage with such musical pieces as *Music of Changes* (1951), in which the musical score is written using chance operations, and even more *HPSCHD* (1969), which uses computer-generated chance operations real-time for various sound sources. Indeed the technology available nowadays makes it possible for a more flexible and complex interactions.

7. INTERACTION

In the field of digital arts, the concept of interaction has repeatedly been identified as a fundamental characteristic and this notion is one of the fundamental advance brought by the use of computers in art installations [9, 5, 11]. In the paradigm of the interface, interaction has often been considered as a necessarily reduced and incomplete dialogue between the human and the machine, as a means of access that will always be frustrating since it is imperfectly codified according to the digital reality hidden at the core of the machine. In taking up this route, numerous artists have done their utmost to devise interfaces that are more or less natural to allow for an improved interaction with their digital works, as if the viewer's immersion should necessarily go through a complex technological apparatus. That approach, however, forgets that immersion is cognitive before being perceptive, the *reality* of a work clearly being invented and recreated by the viewer and not just perceived and undergone. A few artists have designed sound installations linking perception and imagination.

A pioneering digital artwork using contextual information and real-time generated music is *Ping* by Chris Chafe and Greg Niemeyer, first presented in the *Art and Technology* exhibition at San Francisco MOMA in 2001. The title refers to the *ping* command in Unix, which can measure the time needed to reach any specific TCP/IP address on the internet. Therefore in this work, several rhythms are created by several distinct computers, each one *pinging* a specific TCP/IP address. This environment could hardly be defined as musical but somehow aimed at representing in an auditory manner some (partial) topology of the internet, resulting in an intuitive notion of *close* and *distant* in this virtual network. Another more recent example is *Emotional Traffic (e-traffic)* by Maurice Benayoun, exhibited at V2, Rotterdam, in 2005. This work uses the measurement of the number of occurrences of some words representing emotions (like hope, fear, satisfaction, etc) in search engines on the web to produce a sound environment echoing the *emotional state* of the planet at a given instant. These words were also graphically represented within visual representations on projected screens.

In *Red Light Spotters*, we do not want to make explicit the mediation between the ambient context and the generated music. It has to be intuitive and immersive and therefore the fact that this is a situated installation, within an effective landscape that could be visually explored is utmost

⁶<http://www.ableton.com/>

⁷<http://www.native-instruments.com>

⁸<http://opensoundcontrol.org/>

important. The play between the natural visual stimuli and the artificial, computer-generated audio stimuli are one of the key features of this work, but they might remain fuzzy in the mind of the spectators. We do not aim at having, for instance, the spectators identifying each particular blinking light and its association to a precise sound. Nor do we need to explain to the spectators the details of the tempo-tracking machinery or of the light-grouping associations. This is more a *global* interaction: the nebula of the flashing lights creates a nebula of sounds and one can find an intuitive link between them, but there is no need to further explicit the precise links. If willing to do so, the spectator can use his eyes and ears to track down any particular relationship. In *Red Light Spotters*, the interaction that is created consists of a multi-sensory relationship between an external context, the urban landscape offered to the eyes of the spectators, and an internal exhibition space filled with rhythmic music, mediated through tempo-tracking, sound samples and audio-effects associations.

8. CONCLUSIONS

Red Light Spotters is an interactive sound installation that turns an ambient dynamic urban landscape into real-time generated musical soundscape. A key-point is the extraction of tempo in real-time for the groups of blinking lights that form rhythm generators, as this symbolic information can be used in many ways: not only to control the actual tempo of a musical instrument but to influence any parameter in the real-time sound generation (e.g. the waveform of an oscillator, the reverb delay, etc).

Another interesting point is that this installation is very flexible: it can either creates a sound environment in a purely *automatic* mode, or allow for a real-time performance by a musician or sound artist who could modify in real-time the grouping of blinking lights and the association between those rhythm generators and sounds. It can also make room for a mixed approach, where a musician blend his own music with the *Red Light Spotters*.

In its current stage, this installation is closely linked to the Tokyo urban landscape (Fig.7) and thus to be shown during night time in some Tokyo city view observatory, but could be further developed and adapted to other cities and urban landscape, as the only thing is to have some rhythmic flow that has to be captured by the camera sensors and analyzed by the tempo-tracking algorithm. It can be for instance the flow of people or cars in a particular urban location.



Figure 7: Tokyo's Red Light Spots

9. REFERENCES

- [1] S. Arom. *Polyphonies et polyrythmies instrumentales d'afrique centrale, structure et méthodologies*. Sela, 1985.
- [2] P. Boulez. *Penser la musique aujourd'hui*. Gonthier médiations, 1963.
- [3] N. Bourriaud. *Relational Aesthetics*. Les Presse Du Reel, Paris, France, 1998.
- [4] N. Collins. *Beat Induction and Rhythm Analysis for Live Audio Processing*. 2004.
- [5] E. Couchot. *La technologie dans l'art: de la photographie a la realite virtuelle*. Jacqueline Chambon, Paris, France, 1998.
- [6] D. Huron. *Sweet Anticipation: Music and the Psychology of Expectation*. in press, 2006.
- [7] A. Klapuri and M. Davy. *Signal Processing Methods for Music Transcription*. Springer US, 2006.
- [8] T. H. A. Kristoffer Jensen. Real-time beat estimation using feature extraction. In *Computer Music Modeling and Retrieval*, volume 2771/2004, pages 13–22. Springer Berlin / Heidelberg, 2004.
- [9] C. Paul. *Digital Art*. Thames and Hudson, London, U.K., 2003.
- [10] J.-M. Pelletier. A shape-based approach to computer vision musical performance systems. In *NIME*, pages 197–198, 2004.
- [11] F. Popper. *from Technological to Virtual Art*. MIT Press, Cambridge, USA, 2007.
- [12] L. Rabiner. A tutorial on hidden markov models and selected applications in speech. *Proceeding of the IEEE*, 77:257–286, 1989.
- [13] J. Seibt. Free process theory: Towards a typology of occurrences. *Axiomathes*, 14:23–55, 2004.
- [14] J. Seppanen. Tatum grid analysis of musical signals. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.
- [15] B. Simeone. Optimal grid search. private communications, 2006.
- [16] S.-W. L. Tae-Yong Kim, Jihun Park. Object boundary edge selection for accurate contour tracking using multi-level canny edges. In *Image Analysis and Recognition*, volume 3212/2004 of *Lecture Notes in Computer Science*, pages 536–543. Springer Berlin / Heidelberg, 2004.

Bibliographie

- ISMIR 2003, 4th International Conference on Music Information Retrieval, Baltimore, Maryland, USA, October 27-30, 2003, Proceedings*, 2003. [177](#), [183](#), [184](#)
- S. N. Afriat. The output limit function in general and convex programming and the theory of production. *Econometrica*, 39 :309–339, 1971. [53](#)
- S. N. Afriat. A theorem on shadow prices. *Econometrica*, 41 :1197–1199, 1973. [53](#)
- P. E. Allen and R. B. Dannenberg. Tracking musical beats in real time. In *In Proc. of the 1990 Intl. Computer Music Conf*, pages 140–143, 1990. [23](#)
- M. Alonso, G. Richard, and B. David. Tempo estimation for audio recordings. *Journal of New Music Research*, 36(1) :17–25, 2007. [23](#)
- V. Arifi, M. Clausen, F. Kurth, and M. Müller. Automatic synchronization of music data in score-midi- and pcm-format. In *ISMIR DBL [2003]*. [5](#)
- S. Arom. *Polyphonies et polyrythmies instrumentales d’Afrique Centrale, structures et méthodologies , Livres 1-4*. Selaï, 1985a. [6](#), [7](#), [14](#), [21](#)
- S. Arom. *Polyphonies et polyrythmies instrumentales d’Afrique Centrale, structures et méthodologies - Livre V : Organisation du temps dans les musiques africaines*. Selaï, 1985b. [8](#)
- Y.-A. Bois and A. Reiter-McIntosh. Piet mondrian, ”new york city”. *Critical Inquiry*, 2(14) : 244–277, 1988. [15](#)
- P. Boulez. *Penser la musique aujourd’hui*. Gonthier médiations, 1963. [9](#)
- J. C. Brown. Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4) :1953–1957, 1993. [23](#)
- A. T. Cemgil. Tempo tracking and rhythm quantization by sequential monte carlo. In *In Proc. NIPS*. MIT Press, 2001. [23](#), [128](#)

- A. T. Cemgil. Monte carlo methods for tempo tracking and rhythm quantization. *Journal of Artificial Intelligence Research*, 18 :45–81, 2003. 128
- A. T. Cemgil. *Bayesian Music Transcription*. PhD thesis, Radboud University of Nijmegen, 2004. URL <http://staff.science.uva.nl/~cemgil/papers/cemgil-thesis.pdf>. 73, 99
- A. T. Cemgil, B. Kappen, and P. Desain. Rhythm quantization for transcription. *Computer Music Journal*, 24(2) :60–76, 2000. ISSN 0148-9267. doi : <http://dx.doi.org/10.1162/014892600559218>. 23
- J. Chadabe. Remarks on computer music culture. *Computer Music Journal*, 24(4) :9–11, Winter 2000. xx, 1
- A. Chakraborty. Markov chain monte carlo 1. examples. *RESONANCE*, 2002a. 105
- A. Chakraborty. Markov chain monte carlo 2. a method of simulation. *RESONANCE*, 2002b. 105
- M. Chemillier. Mathématiques de tradition orale. *Math. Sci. hum*, 45(178) :11–40, 2007. 4
- J. L. Chen, V. B. Penhune, and R. J. Zatorre. Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex Advance Access published*, 2008. 20, 102
- S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4) :327–335, 1995. URL <http://dx.doi.org/10.2307/2684568>. 106
- B. Y. Chua and G. Lu. Improved perceptual tempo detection of music. In *MMM '05 : Proceedings of the 11th International Multimedia Modelling Conference*, pages 316–321, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2164-9. URL <http://dx.doi.org/10.1109/MMMC.2005.49>. xxii, 11, 23, 25
- K. L. Chung. *Elementary Probability Theory With Stochastic Processes*. Springer, 1978. 147
- Y. Cojan. Perception et production de tempo : Etudes comportementale et electrophysiologique. *thèse*, 2006. 9
- S. Coleman, G. Nouno, and T. Suzuki. Rameses 2000. article à paraître, performance au théâtre des bouffes du nord, Paris, 1999. xxi, 3, 14
- N. Collins. Beat induction and rhythm analysis for live audio processing. Technical report, University of Cambridge, 2004. 21
- A. Cont. A coupled audio/tempo model for real-time alignment of polyphonic audio to music score. *Pami*, 2008a. 10, 117

- A. Cont. *Modeling Musical Anticipation : From the time of music to the music of time*. PhD thesis, UCSD IRCAM, 2008b. 10, 19
- S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30 :39–58, 2001. 23
- S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *In Proc. ISMIR*, pages 159–165, 2003. 5
- S. Dixon, W. Goebel, and E. Cambouropoulos. Perceptual smoothness of tempo in expressively performed music. *Music Perception*, 23(3) :195–214, 2004. 17, 25
- A. L. Dontchev, H. Qi, and L. Qi. Convergence of newton’s method for convex best interpolation, 1999. 93
- J. S. Downie. The scientific evaluation of music information retrieval systems : Foundations and future. *Comput. Music J.*, 28(2) :12–23, 2004. ISSN 0148-9267. URL <http://dx.doi.org/10.1162/014892604323112211>. 16
- C. Drake and A. Penel. Sources of timing variation in music performances. *Psychological Research*, 61 :12–32, 1998. 75
- C. Drake, M. R. Jones, and C. Baruch. The development of rhythmic attending in auditory sequences : attunement, referent period, focal attending. *Cognition* 77, pages 251–288, 2000. 20, 21
- S. Dubnov. Spectral anticipations. *Comput. Music J.*, 30(2) :63–83, 2006. ISSN 0148-9267. URL <http://dx.doi.org/10.1162/comj.2006.30.2.63>. 10, 19
- R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000. 18
- D. Eck. A positive-evidence model for rhythmical beat induction. *Journal of New Music Research*, 30(2) :187–200, 2001a. 23
- D. Eck. A network of relaxation oscillators that finds downbeats in rhythms. In *Artificial Neural Networks – ICANN 2001 Proceedings*, pages 1239–1247. Springer, 2001b. 23
- D. P. Ellis. Beat tracking by dynamic programming. Technical report, LabROSA Columbia University, 2007. 23
- J. Foote and S. Uchihashi. The beat spectrum : A new approach to rhythm analysis. *icme*, 00 : 224, 2001. URL <http://doi.ieeecomputersociety.org/10.1109/ICME.2001.1237863>. 23

- G. D. Forney. The viterbi algorithm. *Proceedings of The IEEE*, 61(3) :268–278, 1973. 107
- M. Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2) :159–171, 2001. 23
- F. Gouyon and P. Herrera. Exploration of techniques for automatic labeling of audio drum tracks instruments. In *Proceedings of MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, Spain, 2001. 23
- F. Gouyon, G. Widmer, X. Serra, and A. Flexer. Acoustic cues to beat induction. a machine learning perspective. *Music Perception. UCPress*, 24 :177–188, 2006. URL <http://mtg.upf.edu/files/publications/bc6559-MP-2006.24.2.177-Gouyon.pdf>. 18, 19
- J. A. Grahn and M. Brett. Rhythm and beat perception in motor areas of the brain. *J. Cognitive Neuroscience*, 19(5) :893–906, 2007. ISSN 0898-929X. URL <http://dx.doi.org/10.1162/jocn.2007.19.5.893>. 3, 20
- O. Z. Guerino Mazzola. Tempo curves revisited : Hierarchies of performance fields. *Computer Music Journal*, 18 :40–52, 1994. xxii, 11, 69, 70, 71, 127
- S. W. Hainsworth. *Techniques for the Automated Analysis of Musical Audio*. PhD thesis, University of Cambridge, UK, September 2004. URL <http://citeseer.ist.psu.edu/cache/papers/cs/31304/http:zSzzSzwww-sigproc.eng.cam.ac.ukzSz~swH21zSzthesis.pdf/hainsworth03techniques.pdf>. 21
- S. W. Hainsworth and M. D. Macleod. Particle filtering applied to musical tempo tracking. *EURASIP J. Appl. Signal Process.*, 2004(1) :2385–2395, 2004. ISSN 1110-8657. URL <http://dx.doi.org/10.1155/S1110865704408099>. 120
- J. Harvey, G. Nouno, and A. Cont. Speakings. 2008. 117
- A. Holzapfel and Y. Stylianou. Beat tracking using group delay based onset detection. *ISMIR*, 2008. 23
- G. T. HOOPEN, T. SASAKI, Y. NAKAJIMA, G. REMIJN, B. MASSIER, K. S. RHEBERGEN, and W. HOLLEMAN. Time-shrinking and categorical temporal ratio perception : Evidence for a 1 :1 temporal category. *Music Perception*, 24(1), 2006. 17
- D. Huron. *Sweet Anticipation : Music and the Psychology of Expectation*. in press, 2006. 10
- M. Imberty. *La Musique Creuse Le Temps*. Unknown, 2005. 8
- F. V. Jensen. *An introduction to bayesian networks*. London , UCL Press, 1996. 157
- F. V. Jensen. *Bayesian Networks and Decision Graphs*. Springer, 2002. 157

- K. Jensen and T. H. Andersen. Beat estimation on the beat. In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 87–90, 2003. [23](#)
- M. R. Jones. Time, our lost dimension : toward a new theory of perception, attention, and memory. *Psychological Review*, pages 323–355, 1976. [20](#)
- A. P. Klapuri. Musical meter estimation and music transcription. In *In Proc. Cambridge Music Processing Colloquium*, pages 40–45, 2003. [23](#)
- E. W. Large. Periodicity, pattern formation, and metric structure. *Journal of New Music Research*, 30(2) :173–185, 2001. [119](#)
- E. W. Large. *Dynamic representation of musical structure*. PhD thesis, Ohio State University, 1994. [14](#), [113](#)
- E. W. Large. Beat tracking with a nonlinear oscillator. *IJCAI-95 Workshop on Artificial Intelligence and Music*, pages 24–31, 1995. [23](#), [117](#)
- E. W. Large and M. R. Jones. The dynamics of attending : How people track time-varying events. *Psychological Review*, 106(1) :119–159, 1999. [xxii](#), [xxiii](#), [11](#), [13](#), [21](#), [102](#), [110](#), [112](#), [120](#), [128](#)
- E. W. Large and J. F. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6(2,3) :177 – 208, 1994. [21](#)
- E. W. Large and C. Palmer. Perceiving temporal regularity in music. *Cognitive Science*, (26) :1–37, 2002. [xxii](#), [xxvii](#), [11](#), [13](#), [14](#), [102](#), [113](#), [119](#), [127](#)
- J. Laroche. Estimating tempo, swing and beat locations in audio recordings. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 135–138, 2001. [23](#)
- Y. Levin, Y. Levin, A. Ben-israel, and A. Ben-israel. The newton bracketing method for convex minimization. *Computational Optimization and Applications*, 2000. [93](#)
- G. Lewis. Too Many Notes : Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, 10, 2000. [14](#)
- J. Loach. Le corbusier and the creative use of mathematics. *British Journal for the History of Science*, 31 :185–215, 1998. [15](#)
- Longuet-Higgins. The perception of music. In *Proceedings of the Royal Society of London.*, 1979. [23](#)

- O. Luening. An unfinished history of electronic music. *Music Educators Journal*, 55 :42–49+135–142+145, 1968. [xx](#), [1](#)
- B. Maderna and N. Verzina. *Bruno Maderna, Etude Historique Et Critique*. Harmattan, 2003. [8](#)
- P. S. Maybeck. *Stochastic models, estimation, and control*, volume 141 of *Mathematics in Science and Engineering*. Academic Press, 1979. [104](#), [117](#), [120](#), [128](#)
- J. D. McAuley. *Perception of time as phase : toward an adaptive-oscillator model of rhythmic pattern processing*. PhD thesis, Indianapolis, IN, USA, 1995. [23](#)
- M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri. Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research*, 36(1) : 1–16, 2007. [21](#)
- M. Mezzadri, B. van der Werf, and Octurn. Xp's. 2003-2007. [xxi](#), [3](#)
- I. Mierswa and K. Morik. Automatic feature extraction for classifying audio data. *Machine Learning*, 58(2-3) :127–149, 2005. URL <http://springerlink.metapress.com/index/10.1007/s10994-005-5824-7>. [18](#)
- T. Miura and I. Shioya. Similarity among melodies for music information retrieval. In *CIKM '03 : Proceedings of the twelfth international conference on Information and knowledge management*, pages 61–68, New York, NY, USA, 2003. ACM. ISBN 1-58113-723-0. URL <http://doi.acm.org/10.1145/956863.956877>. [5](#)
- K. P. Murphy. *Dynamic bayesian networks : representation, inference and learning*. PhD thesis, UC Berkeley, 2002. URL <http://www.cs.ubc.ca/~murphyk/Thesis/thesis.html>. Chair-Stuart Russell. [104](#), [128](#)
- L. Nono. *Ecrits*. Christian Bourgois Editeur, 1993. [8](#)
- M. Olivetti-Belardinelli. Ricerca sui fattori che determinano la preferenza per l'esecuzione ritmica umana o computerizzata. *Comunicazioni scientifiche di psicologia generale*, 10 : 169–192, 1993. [8](#)
- M. Olivetti-Belardinelli and S. Nicole. Componenti ritmiche in serie di atti motorisé semplici : per una nuova modalità di analisi. *Comunicazioni scientifiche di psicologia generale*, 10 : 59–90, 1993. [9](#)
- M. Olivetti-Belardinelli and E. Pessa. Chaotic deterministic dynamic in rhythmic structures of spontaneous musical activity. *Comunicazioni scientifiche di psicologia generale*, 5 : 103–118, 1991. [9](#)

- M. Olivetti-Belardinelli, D. Miglio, C. Reali, and G. P. S. l'influenza del ritmo spontaneo nella elaborazione modulare della sagoma del corpo umano. *Comunicazioni scientifiche di psicologia generale*, 10 :31–42, 1993. 9
- G. Paczynski. *Une histoire de la batterie de jazz*. Outre Mesure, 2000. 19
- R. Parncutt. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, 11(4) :409–464, 1994. 23
- M. Parry and I. Essa. Rhythmic similarity through elaboration. In *ISMIR DBL [2003]*. 5
- J. Paulus and A. Klapuri. Measuring the similarity of rhythmic patterns. In M. Fingerhut, editor, *Proc. of the Third International Conference on Music Information Retrieval*, pages 150–156, Paris, Oct. 2002. ISBN 2-84426-166-3. 23
- J. Pearl. *Causality : Models, Reasoning, and Inference*. Selaf, 1985. 10
- J. Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Morgan Kaufmann publishers, inc, 1988. 157
- G. Peeters. Template-based estimation of time-varying tempo. *EURASIP J. Appl. Signal Process.*, 2007(1) :158–158, 2007. ISSN 1110-8657. URL <http://dx.doi.org/10.1155/2007/67215>. 23
- M. Pfeuty, R. Agot, and V. Poutha. Processes involved in tempo perception : A cnvanalysis. *Psychophysiology*,, 40 :69–76, 2003. 101, 102
- M. Puckette. Max at seventeen. *Comput. Music J.*, 26(4) :31–43, 2002. ISSN 0148-9267. URL <http://dx.doi.org/10.1162/014892602320991356>. xx, 1
- L. Rabiner. A tutorial on hidden markov models and selected applications in speech. *Proceeding of the IEEE*, 77 :257–286, 1989. 107, 108
- C. Raphael. Automated rhythm transcription. In *In Proc. Int. Symposium on Music Inform. Retrieval. (ISMIR)*, pages 99–107, 2001. 23
- D. G. Regina. The 6th international conference on music information retrieval. *Computer Music Journal*, 30(2) :90–92, 2006. 16
- D. Rosenthal. Emulation of human rhythm perception. 16(1) :64–76, Spring 1992. 23
- M. Sadakata, P. Desain, and H. Honing. The bayesian way to relate rhythm perception and production. *Music Perception*, 23 :269–288, 2005. 102, 106

- E. D. Scheirer. Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1) :588–601, 1998. doi : <http://dx.doi.org/10.1121/1.421129>. URL <http://dx.doi.org/10.1121/1.421129>. 23
- L. Seebeck, R. M. Kim, and S. M. Kaplan. Emergent temporal behaviour and collaborative work. In *ECSCW*, pages 123–142, 2005. URL <http://www.ecscw.org/2005/paper07.pdf>. 4
- J. Seppanen. Tatum grid analysis of musical signals. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 131–134, 2001. 23, 27, 35, 67
- Sethares and Morris. Beat tracking of musical performances using low-level audio features. *IEE transaction on speech and audio processing*, 13(2), 2005. 23
- B. Simeone. recherche de grille optimale. *communication personnelle*, 2006. 28
- L. M. Smith. *A Multiresolution Time-Frequency Analysis and Interpretation of Musical Rhythm*. PhD thesis, University of Western Australia, Australia, October 2000. 23, 67
- L. M. Smith and H. Honing. Time–frequency representation of musical rhythm by continuous wavelets. *Journal of Mathematics and Music*, 2(2) :81–97, 2008. 67
- M. Stroppa and A. Cont. little i. 2008. 117
- H. Takeda, T. Nishimoto, and S. Sagayama. Automatic rhythm transcription from multi-phonetic midi signals. In *ISMIR DBL [2003]*. 5
- D. Temperley. An evaluation system for metrical models. *Comput. Music J.*, 28(3) :28–44, 2004. ISSN 0148-9267. URL <http://dx.doi.org/10.1162/0148926041790621>. 21
- D. Temperley and D. Sleator. Modeling meter and harmony : A preference-rule approach. *Comput. Music J.*, 23(1) :10–27, 1999. ISSN 0148-9267. URL <http://dx.doi.org/10.1162/014892699559616>. 23
- L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, pages 1701–1728, 1994. 106
- P. Toiviainen. An interactive midi accompanist. *Computer Music Journal*, 22(4) :63–75, 1998. 23
- P. V. Turner. Frank lloyd wright and the young le corbusier. *The Journal of the Society of Architectural Historians*, 42(4) :350–359, 1983. 15
- G. Tzanetakis. Musical genre classification of audio signals. In *IEEE Transactions on Speech and Audio Processing*, pages 293–302, 2002. 23

- G. Tzanetakis and A. Kapur. Computational ethnomusicology. *journal of interdisciplinary music studies*, 1(2) :1–24, fall 2007. 4
- C. Uhle, J. Herre, and F. Fraunhofer. Estimation of tempo, micro time and time signature from percussive music. In *In Proc. Digital Audio Effects Workshop DAFx*, 2003. 23
- N. L. Wallin, B. Merker, and S. Brown. *The Origins of Music*. MIT Press, 2000. Ch.17 How music fixed "Nonsense" into significant formulas : on rhythm, repetition and meaning. 15
- Y. Wang and M. Vilelmo. A compressed domain beat detector using mp3 audio bitstreams. In *MULTIMEDIA '01 : Proceedings of the ninth ACM international conference on Multimedia*, pages 194–202, New York, NY, USA, 2001. ACM. ISBN 1-58113-394-4. doi : <http://doi.acm.org/10.1145/500141.500172>. 23
- M. Wells. Third international conference on music information retrieval. *Computer Music Journal*, 27(2) :108–111, 2003. 16
- M. Wright. *Computer-Based Music Theory and Acoustics*. PhD thesis, Stanford University, CA, USA, March 2008. URL <http://ccrma.stanford.edu/~matt/diss/>. 3, 18, 19