

Captation gestuelle pour une percussion augmentée

Charles VERRON

**Mémoire de stage de DEA ATIAM
année 2003-2004**

Université Paris 6

La kitchen
78, avenue de la République
75011 Paris

Sous la responsabilité de Thierry Coduys et Cyrille Henry

Remerciements

Je tiens tout d'abord à remercier Thierry Coduys et Cyrille Henry sans qui ce stage n'aurait pas été possible. Merci aussi à toute l'équipe de La kitchen, pour la bonne humeur, et pour leur aide pendant le stage.

Table des matières

Table des matières	i
Introduction	1
I Le geste instrumental	3
I.1 L'instrument et le geste	3
I.2 Notions d'avant et d'après-geste	3
II Etat de l'art	5
III Chaîne d'acquisition et de traitement du geste	7
III.1 Les capteurs	7
III.2 Les interfaces gestuelles	8
III.2.1 Le Toaster	8
III.2.2 La Rimas Box	8
III.3 Les logiciels de traitement et de synthèse sonore	9
IV Captation du geste du percussionniste	11
IV.1 Analyse des gestes du percussionniste	11
IV.1.1 Modèle du geste	11
IV.1.2 Extraction des paramètres pertinents	12
IV.2 Mesures	13
IV.3 Choix des capteurs et de l'interface	16
IV.3.1 Capteurs	16
IV.3.2 Interface	18
V Données de la Rimas Box et prétraitement	21
V.1 Format du signal de sortie de la Rimas Box	21
V.2 Prétraitement : correction du signal de sortie de la Rimas Box	24
V.2.1 Sous-échantillonnage	24
V.2.2 Filtrage passe bas	24
VI Traitements des données gestuelles	31
VI.1 Détermination de la vitesse angulaire des baguettes	31
VI.2 Détection de l'attaque	33
VI.3 Détermination de la vitesse des coups	40
VI.4 Réalisation en Max/MSP	40

VII Mapping des données pour une synthèse sonore	43
VII.1 Définition	43
VII.2 Application pour la percussion augmentée	44
VIII Perspectives	45
VIII.1 Classification et reconnaissance des frappes	45
VIII.2 Captation d’autres paramètres du mouvement	45
Conclusion	47
Bibliographie	49

Introduction

La kitchen est une entreprise spécialisée dans la captation gestuelle à vocation artistique. Son laboratoire développe essentiellement des systèmes électroniques de captation, réalisés sur mesure pour des oeuvres musicales ou des installations sonores.

Le stage s'inscrit dans un projet réunissant La kitchen et le percussionniste Daniel Champolini, en vue de la création d'une pièce musicale. Le but est de réaliser une percussion augmentée, destinée à être utilisée en concert. Parallèlement, un travail de recherche sur les interfaces de La kitchen a été réalisé, permettant de spécifier les limites des systèmes de captation actuels et les axes de développement futurs.

Cette introduction se propose de donner une brève description du système réalisé puis d'annoncer le plan du rapport.

Notre percussion augmentée est un système de captation du geste intégré sur des baguettes de percussion traditionnelles. Il s'agit d'un ensemble de capteurs analogiques d'accélération et d'angle, disposés pour recueillir des paramètres pertinents du geste de l'instrumentiste. Un traitement avec des logiciels temps réel comme Pure-Data ou Max/MSP réalise une synthèse sonore en fonction du mouvement, de manière à enrichir le discours musical.

Le système a été nommé « percussions augmentées », puisque sa fonction est avant tout de compléter le son acoustique naturellement émis par l'instrument, tout en apportant un moyen d'expression gestuel nouveau. Les baguettes peuvent néanmoins constituer une véritable « percussion virtuelle » dès lors qu'on les utilise sur une surface qui étouffe le son naturel des coups portés.

Le rapport compte huit parties. La première est une étude préliminaire sur le geste instrumental. Dans le deuxième chapitre, nous décrivons les systèmes de percussions virtuelles déjà existants. La troisième partie décrit une chaîne d'acquisition gestuelle classique. Enfin les chapitres IV, V, VI et VII décrivent le travail réalisé pendant le stage, et la réalisation des percussions augmentées. La dernière partie évoque les possibilités d'évolution du système.

Chapitre I

Le geste instrumental

Nous présentons ici une étude succincte des liens qui unissent le geste, l'instrument et la musique. La première partie donne quelques définitions préalables. La deuxième partie apporte une segmentation chronologique du geste instrumental adaptée pour notre application des percussions augmentées.

I.1 L'instrument et le geste

Nous donnons ici simplement quelques définitions. Pour une étude plus approfondie, nous invitons le lecteur à examiner les nombreux écrits de Claude Cadoz (ex : [3], [4], [5]) et de Marcello Wanderley (ex : [14], [15]) sur le sujet.

- On appelle **geste** tout mouvement ou posture du corps ou d'une partie du corps [2].
- **Le geste instrumental** est un geste exécuté en contact avec une matière (dans notre cas un instrument de musique), par opposition au geste à main nue.
- **L'instrument de musique** est le corps qui effectue la transformation du geste au son, dans le temps même des actions.
- **Les gestes de l'instrumentiste** sont l'ensemble des actions physiques effectuées en situation de jeu musical.

I.2 Notions d'avant et d'après-geste

Cadoz distingue trois classes de gestes instrumentaux, qui concernent les gestes du musicien participant à la production du son :

- les gestes d'excitation, qui fournissent l'énergie à l'instrument (souffler dans l'anche d'une clarinette)
- les gestes de modification, qui modifient la chaîne instrumentale principale (déplacer sa main sur le manche d'un violon)
- les gestes de sélection, qui sélectionnent un sous-ensemble de chaînes instrumentales (appuyer sur une touche de piano)

Pour notre application de la percussion augmentée, nous avons été amenés à considérer surtout des gestes de l'instrumentiste n'ayant pas une incidence directement audible. En effet, si les

mouvements du percussionniste sont riches et continus dans le temps, l'instrument a la particularité de ne produire du son qu'à des instants très précis, aux moments des impacts de la baguette.

Cela nous a conduit à introduire les notions d'avant et d'après-geste, pour désigner les gestes instrumentaux qui ne sont pas concomittents à l'émission du son.

L'avant-geste regroupe l'ensemble des mouvements effectués par l'instrumentiste avant l'émission sonore. Cette phase, qui s'exécute consciemment ou non, a pour but de préparer l'action qui génère effectivement le son.

Le geste se limite, d'après notre segmentation, à la partie du geste instrumental qui produit concrètement le phénomène sonore audible.

L'après-geste est constitué des mouvements qui viennent après l'émission sonore. Il peut s'agir selon les cas d'une phase de relâchement musculaire, d'un mouvement dû à la réaction physique de l'instrument, ou de la continuation du geste qui a généré le son.

Nous segmentons ainsi les gestes de l'instrumentiste en trois phases chronologiques :

- La phase de préparation : « l'avant-geste »
- La phase génératrice de son : « le geste »
- La phase de relâchement ou de continuation : « l'après-geste »

Notons dès à présent que les dénominations choisies pour les trois phases constituent un abus de langage certain, puisque l'avant et l'après-geste font en réalité parties intégrantes du geste instrumental. Cependant nous avons choisi ces termes pour bien montrer l'enchaînement chronologique qui relit l'avant-geste, le son, et l'après-geste.

Examinons l'exemple de la percussion en considérant une frappe simple comme le geste instrumental dans son ensemble.

La main qui tient la baguette est dans une position basse de repos. Dans un premier temps, le percussionniste élève la baguette à une certaine distance de l'instrument. Il l'abaisse ensuite en lui donnant de la vitesse. Toute cette phase constitue l'avant-geste.

D'après notre segmentation qui connaît ici un cas limite, le geste du percussionniste se réduit à l'instant de l'impact, lors duquel la baguette transmet son énergie cinétique à l'instrument, ce qui génère le son.

Enfin, Le percussionniste peut choisir de relâcher son action, abandonnant la baguette à la réaction naturelle sur l'instrument, ou de remonter la main en position haute. Cette dernière phase constitue l'après-geste.

Considérons maintenant l'exemple d'un instrument à corde comme le violon. La main qui tient l'archet est souvent en mouvement, sans pour autant que ce dernier soit en train de frotter les cordes pour générer du son. Selon les modes de jeu, l'archet doit attaquer les cordes avec une certaine vitesse initiale, qui lui est conférée par l'instrumentiste dans une phase d'avant-geste. De même, l'archet peut continuer ou non son mouvement après avoir frotté les cordes, phase qui constitue l'après-geste.

Chapitre II

Etat de l'art

Très peu de travaux ont été réalisés sur l'étude et la modélisation des gestes du percussionniste. Plusieurs systèmes de percussions virtuelles ont néanmoins été développés. La plupart d'entre eux ne s'intéressent qu'à la modélisation de l'impact pour le déclenchement d'évènements sonores. D'autres réalisations, comme celle du Scrim de Bordeaux, captent uniquement les gestes du percussionniste dans le vide, et ne s'intéressent pas aux impacts réels. Le système de percussions augmentées que nous envisageons se place donc à la frontière entre toutes les réalisations actuelles.

Le radio bâton Il s'agit du premier système capable de suivre la position de deux baguettes dans les trois directions de l'espace. Max Matthews l'a réalisé pour le déclenchement par le geste d'évènements musicaux [10]. Cependant, ce système n'a pas été conçu pour capter les gestes d'un percussionniste en situation de jeu réel, et son but n'est pas d'obtenir une percussion virtuelle à proprement parler.

La percussion virtuelle du Scrim Le Scrim de Bordeaux développe actuellement un projet de percussion virtuel [9] qui collecte des informations sur les gestes d'un percussionniste dans le vide. Ce système est basé sur l'outil de tracking « Flock of Bird », qui permet d'obtenir la position des baguettes en trois dimensions pour le suivi des gestes du percussionniste. Ce type de matériel semble néanmoins présenter quelques inconvénients pour notre application. La fréquence d'échantillonnage de 100 Hz est trop basse pour détecter un impact sur une surface réelle.

La percussion virtuelle du Gmem Le Gmem de Marseille a également développé un système de percussion virtuelle (www.gmem.org). Ici, le geste est capté à l'aide d'une caméra. Le champ de la caméra est divisé en plusieurs zones situées autour du musicien. Un traitement d'image réalisé avec la librairie Soft VNS pour Max permet de détecter un mouvement dans l'une de ces zones, pour déclencher un son particulier. Compte tenu de la fréquence d'échantillonnage très basse (15 à 30 Hz) du système optique, cette application n'est pas destinée à capter le geste instrumental naturel d'un percussionniste. Le musicien doit s'adapter à l'instrument.

Les batteries électroniques Plusieurs grandes marques d'instruments de musique ont développé des batteries virtuelles. Le musicien joue sur un ensemble de « pads » qui communiquent via MIDI avec un générateur de sons. Ces systèmes présentent néanmoins quelques inconvénients. Tout d'abord, compte tenu de la norme MIDI, les valeurs sont codées sur 7 bits seulement. Mais surtout,

ces systèmes permettent uniquement de détecter des frappes, et non de collecter des informations sur l'avant et l'après-geste de l'instrumentiste.

Enfin Daniel Champolini, en temps que percussionniste professionnel, est insatisfait de la latence des batteries électroniques actuelles. Une oreille très entraînée ressent les millisecondes qui séparent la frappe et le son acoustique du son synthétique. Ce décalage est très gênant et même rédibitoire lors d'une situation de jeu réelle.

Chapitre III

Chaîne d'acquisition et de traitement du geste

La chaîne d'acquisition et de traitement des données gestuelles est la suivante. Nous disposons de plusieurs types de capteurs analogiques (accélération, angle, flexion, pression, lumière, etc.) Ces capteurs sont reliés à une interface de captation gestuelle, qui numérise leur information et la transmet en temps réel à un ordinateur. Les données sont traitées, toujours en temps réel, grâce à des logiciels tels que Max/MSP ou Pure-Data. La synthèse sonore résultant de ces traitements est diffusée à travers une carte son et un système de restitution classiques.

III.1 Les capteurs

Nous avons de nombreux types de capteurs analogiques à notre disposition durant le stage. On peut voir sur la figure III.1 les principaux types, avec de gauche à droite, un capteur d'angle magnétique deux axes, un capteur photométrique, un capteur de pression, un gyroscope, un accéléromètre, un capteur de flexion et un potentiomètre.



FIGURE III.1: Différents capteurs pour le geste

Tous sont de petite taille, avec des dimensions inférieures au centimètre, ce qui permet de les fixer sur une baguette de percussion sans la déséquilibrer. Le choix et la disposition des capteurs pour notre application est discuté au paragraphe IV.3.1.

III.2 Les interfaces gestuelles

Plusieurs modèles d'interface pour la captation gestuelle sont actuellement disponibles. On peut notamment citer le Toaster et le Kroonde, réalisés à La kitchen par Cyrille Henry [6], l'Eo-body de l'Ircam et Eowave [8], la Rimas Box développée au CNMAT [1] et l'Ethersens, sortie depuis peu des laboratoires de l'Ircam. Ces interfaces, dites « sensibles », sont destinées à recueillir des données gestuelles issues de capteurs, et à les transmettre à un ordinateur. Elles utilisent différents protocoles de communication tels que UDP, MIDI, OSC, et présentent différentes résolutions et fréquences d'échantillonnage. Dans le cadre du stage, nous disposons de deux de ces interfaces : le Toaster et la Rimas Box.

III.2.1 Le Toaster

Développé à La kitchen, le Toaster permet de gérer 16 entrées simultanément [6]. Il échantillonne les informations analogiques des capteurs à 200 Hz et les envoie en haut débit via le protocole Open Sound Control (OSC). Les données sont codées sur 16 bits. La figure III.2 montre les faces avant et arrière du Toaster.



(a) vue avant



(b) vue arrière

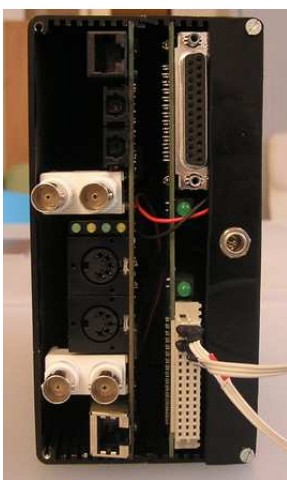
FIGURE III.2: Le Toaster de La kitchen

III.2.2 La Rimas Box

La Rimas Box du CNMAT est à la fois une carte son et une interface gestuelle [1]. Elle est dotée de 26 entrées dont 10 audio et 16 pour les capteurs, et de 10 sorties audio. La fréquence d'échantillonnage est de 5512.5 Hz pour chaque capteur, puisque 8 capteurs se partagent un convertisseur analogique/numérique à 44100 Hz. Les données, codées sur 24 bits, sont envoyées via UDP à un ordinateur sous la forme d'un flux audio échantillonné à 44100 Hz. La figure III.3 montre la Rimas Box que nous avons utilisée.



(a) vue de côté



(b) vue arrière

FIGURE III.3: La Rimas Box du CNMAT

III.3 Les logiciels de traitement et de synthèse sonore

L'application des percussions virtuelles et de la baguette augmentée nécessite plus que jamais de respecter les contraintes du temps réel. La détection d'attaque et la synthèse sonore qui en découle doivent être réalisées dans les millisecondes qui suivent le choc de la baguette contre un objet. Le son virtuel généré et le son acoustique naturel de l'objet percuté doivent être perçus dans le même temps par le musicien pour que le jeu instrumental soit possible. Ces contraintes nous ont amenés à choisir les logiciels Pure-Data [13] et Max/MSP [16] pour réaliser le traitement des données gestuelles et la synthèse sonore.

Pure-Data et Max/MSP sont deux environnements de programmation graphique, dédiés à des applications temps réel et interactives. Ces logiciels se ressemblent beaucoup, car tous deux

proviennent des travaux de Miller Puckette, initiés à l'Ircam à la fin des années 80. Pour plus d'information sur l'origine de ces logiciels, nous invitons le lecteur à consulter la page "<http://free-software.ircam.fr/article.php3>" et les articles [11] et [12].

Nous avons essayé de concevoir des traitements compatibles dans la mesure du possible pour les deux environnements.

Chapitre IV

Captation du geste du percussionniste

Nous avons vu dans le chapitre précédent que nous disposons d'une multitude de capteurs pour relever les gestes du percussionniste. Le choix de la nature et de la disposition de ces capteurs est très important pour notre application. Pour que notre percussion augmentée soit utilisable en situation de jeu réel, les capteurs ne doivent pas gêner l'instrumentiste, et l'équilibre des baguettes doit être respecté. Cependant, les capteurs doivent permettre de collecter un maximum d'informations sur le jeu du percussionniste.

La première étape consiste donc à analyser les gestes du percussionniste, et à en extraire les paramètres pertinents, au sens où ils donnent un maximum d'informations sur son jeu (cf. § IV.1). Après avoir choisi les données du geste qu'il faut capter, nous discuterons des capteurs à utiliser et de leur positionnement. Nous avons effectué une série de mesures préalables en situation de jeu, avec plusieurs capteurs, le but étant de ne garder que ceux qui apportent une information réellement exploitable (cf. § IV.2). Enfin, nous décrirons le dispositif de captation finalement retenu pour notre application (cf. § IV.3).

IV.1 Analyse des gestes du percussionniste

IV.1.1 Modèle du geste

Précisons tout d'abord que dans notre étude nous assimilons les gestes du percussionniste aux mouvements de la baguette de percussion. En situation de jeu réel, ce mouvement peut s'avérer assez complexe. Nous pouvons faire une première approximation et envisager qu'il s'effectue uniquement autour de deux liaisons pivot, constituées par le coude et le poignet du percussionniste.

Une approximation toujours satisfaisante dans un premier temps consiste à simplement considérer le mouvement circulaire de la baguette autour d'un axe situé entre le coude et le poignet. On réduit ainsi le problème à une dimension : l'angle θ entre la baguette et le plan horizontal (cf. figure VI.4).

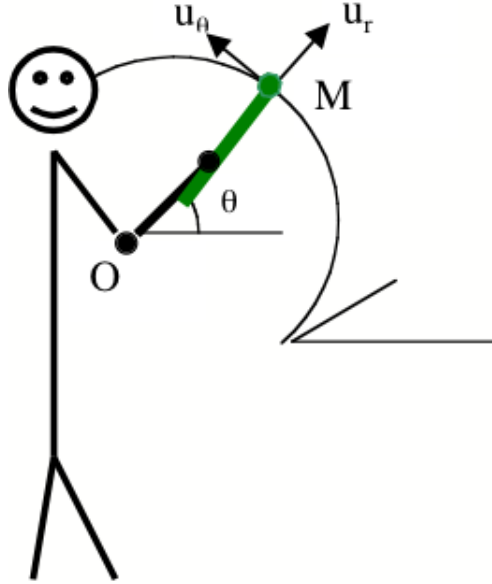


FIGURE IV.1: Modèle unidimensionnel du mouvement de la baguette

Un calcul simple nous donne les formules qui décrivent la vitesse et l'accélération du point M, situé à l'extrémité de la baguette.

Soit O le centre de rotation du mouvement de la baguette et r la distance OM :

$$\overrightarrow{OM} = r\vec{u}_r \quad (IV.1)$$

$$\vec{V} = \frac{d\overrightarrow{OM}}{dt} = r\dot{\theta}\vec{u}_\theta \quad (IV.2)$$

$$\vec{A} = \frac{d\vec{V}}{dt} = r\ddot{\theta}\vec{u}_r + r\dot{\theta}\dot{\theta}\vec{u}_\theta \quad (IV.3)$$

IV.1.2 Extraction des paramètres pertinents

Notre système de captation doit recueillir un maximum d'informations sur l'intention musicale du percussionniste. Les paramètres importants du geste que nous souhaitons capter sont :

- L'angle entre la baguette et le plan horizontal.
- La vitesse angulaire de l'extrémité de la baguette.
- L'occurrence des impacts de la baguette sur une surface solide.
- L'intensité des impacts (la vélocité des coups).

L'angle et la vitesse angulaire nous renseignent sur la totalité du geste de l'instrumentiste : aussi bien pendant la phase d'avant geste, que pendant l'après-geste. Le moyen le plus simple pour obtenir l'angle est de poser un capteur magnétique d'angle sur la baguette. La vitesse angulaire peut être obtenue par dérivation des données de ce capteur. On peut aussi recueillir le carré de la vitesse angulaire à l'aide d'un accéléromètre placé selon l'axe tangentiel à la baguette (cf. formule

VI.3). Enfin, on peut capter la vitesse angulaire avec un gyroscope fixé sur la baguette.

Les impacts de la baguette sur une surface solide sont évidemment essentiels dans le jeu de la percussion, puisqu'ils génèrent le son musical. Leur détection, que nous appelons "**détection d'attaque**", doit permettre de déclencher en temps réel des événements sonores. Plusieurs méthodes semblent possibles pour effectuer cette détection. On peut utiliser les données d'un capteur de pression posé sur la surface percutée ou fixé au bout des baguettes. On peut aussi se servir d'un capteur d'accélération placé selon l'axe normal à la baguette. Selon le choix du capteur, les données devront être traitées par un algorithme de détection spécifique.

L'intensité des chocs de la baguette sur une surface solide nous renseigne sur la vitesse des coups portés par l'instrumentiste et constitue là encore un paramètre crucial du jeu. Dans notre application, cette valeur sera utilisée pour contrôler le volume de la synthèse sonore. Pour obtenir cette information, on peut encore penser à un capteur de pression. La vitesse angulaire de la baguette juste avant l'impact peut également nous renseigner sur la vitesse des coups portés par l'instrumentiste.

IV.2 Mesures

Nous avons effectué une série de mesures en situation de jeu avec Daniel Champolini, pour déterminer quels capteurs s'avéraient réellement efficaces pour notre application. Quatre capteurs ont été utilisés pour recueillir les différents paramètres du mouvement de la baguette décrits au paragraphe IV.1 :

- un capteur magnétique d'angle (donne θ)
- un capteur d'accélération tangentielle (donne $r\dot{\theta}^2$)
- un capteur d'accélération normale (donne $r\ddot{\theta}$)
- un capteur de pression (donne l'intensité du choc)

Les trois premiers capteurs sont fixés au centre de la baguette ce qui permet de conserver son équilibre. Le capteur de pression a dû être posé sur la surface percutée, sa forme ne permettant pas de le fixer simplement à l'extrémité de la baguette. Nous n'avons malheureusement pas effectué de mesure avec un gyroscope lors de la séance avec Daniel Champolini. Un tel capteur peut nous donner $\dot{\theta}$ directement. Cependant, ceux dont nous disposonsaturent à 300° par seconde, ce qui ne semble pas suffisant pour capter la vitesse de la baguette.

Daniel Champolini a effectué une série de frappes, avec des intensités et des vitesses de jeu variées. La figure IV.2 montre les données des quatre capteurs pour une frappe, échantillonnées par la Rimas Box à 5512.5 Hz.

Pour les mesures présentées ici, la surface percutée était un petit carré de mousse rigide posé sur une table de bois. Les mesures effectuées sur des surfaces différentes (plus ou moins rigides) ont donné des résultats sensiblement identiques.

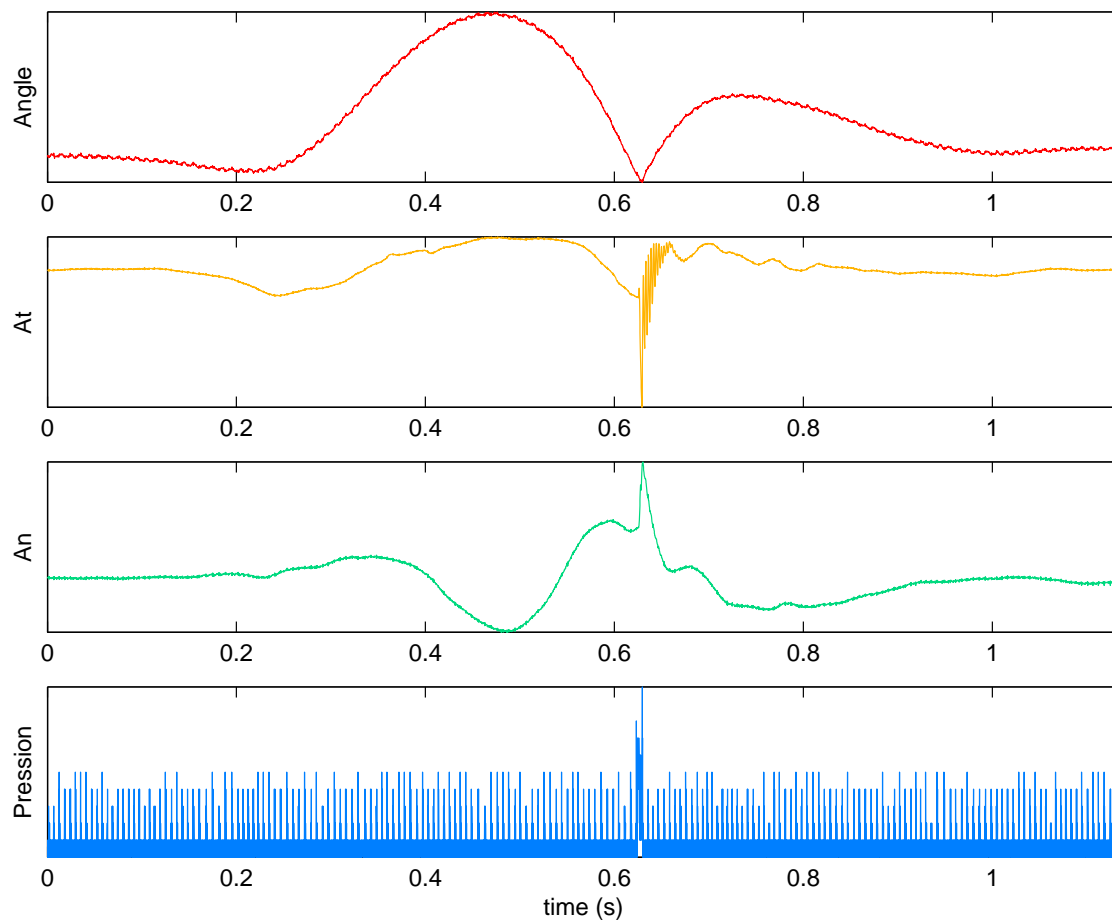
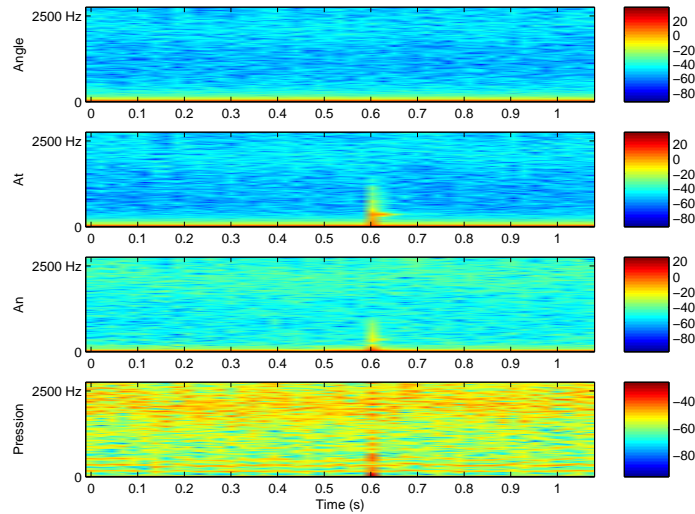


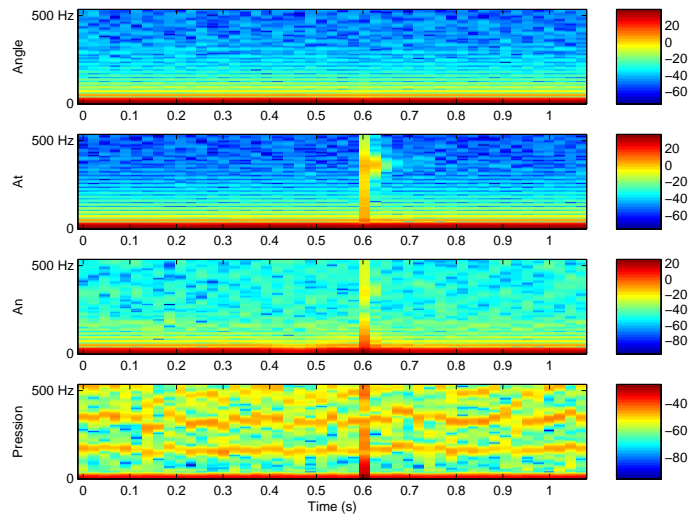
FIGURE IV.2: données temporelles sur une frappe

On remarque sur cet exemple un problème typique du capteur de pression : le signal recueilli est de l'ordre du niveau de bruit, autrement dit le capteur n'a pas "vu" le choc. Ce phénomène est dû à la durée infime de l'impact de la baguette sur le capteur, qui semble souvent inférieure à la milliseconde. Le capteur est incapable de répondre dans des temps aussi courts est la discontinuité du choc n'apparaît pas. La période d'échantillonnage de la Rimas Box étant de l'ordre de 0.2 ms, on arrive également dans les limites de résolution temporelle du système.

La figure IV.3 montre une analyse temps-fréquence des mêmes données. La fenêtre utilisée pour la transformée de Fourier à court terme est une fenêtre de Hanning de 128 points. Chaque transformée de Fourier discrète est calculée sur 1024 points fréquentiels. Le recouvrement temporel entre deux fenêtres successives est de 64 points. On obtient ainsi des résolutions temporelles et fréquentielles satisfaisantes pour l'analyse de nos signaux gestuels.



(a) fréquences de 0 à 2750 Hz



(b) zoom sur les fréquences < 500 Hz

FIGURE IV.3: spectrogrammes sur une frappe

On constate que la quasi-totalité de l'énergie des signaux est concentrée dans les fréquences inférieures à 100 Hz. Seule l'accélération tangentielle présente des hautes fréquences au moment de l'impact de la baguette sur la surface percutée. L'étude des mesures a montré que la fréquence de coupure à -20 dB dans le spectre pouvait alors atteindre 1200 Hz.

Sur la figure IV.4, qui reprend les données temporelles des capteurs d'angle et d'accélération tangentielle, nous spécifions les trois phases importantes du mouvement.

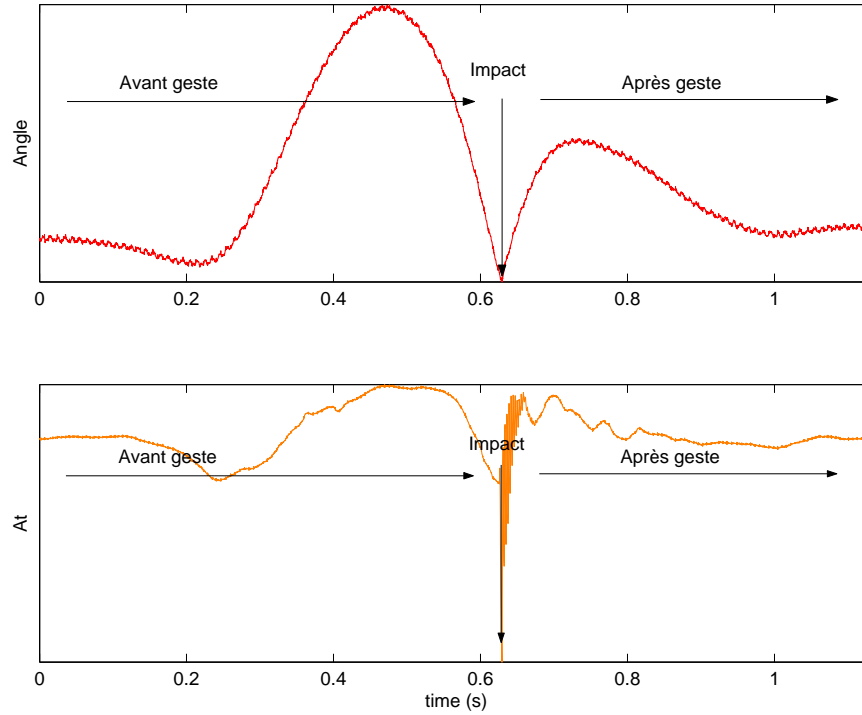


FIGURE IV.4: Les 3 phases du geste du percussionniste

IV.3 Choix des capteurs et de l'interface

IV.3.1 Capteurs

Nous avons précisé au paragraphe IV.1.2 que les paramètres importants du geste que nous souhaitons capter sont l'angle entre la baguette et le plan horizontal, la vitesse angulaire de l'extrémité de la baguette, les impacts de la baguette sur une surface solide et la vitesse des coups.

L'angle entre la baguette et le plan horizontal est donné directement par le capteur magnétique d'angle.

Pour la vitesse angulaire, le capteur d'accélération normale renvoie une valeur proportionnelle à $r\ddot{\theta}$. Il nous renseigne donc bien sur la norme de la vitesse angulaire. Cependant, cette donnée ne permet pas de connaître le sens du mouvement de la baguette, qui s'effectue alternativement vers le haut ou vers le bas. Nous avons donc choisi de déterminer la valeur relative de la vitesse angulaire par dérivation du capteur magnétique d'angle, plutôt que d'utiliser un capteur d'accélération normale. Cette opération a aussi l'avantage de réduire le nombre de capteurs embarqués sur la baguette.

Les mesures décrites au paragraphe IV.2 ont montré que notre capteur de pression est inefficace pour détecter l'instant précis des impacts et leur intensité. Cette solution était de toute façon compromise, pour des raisons de résistance du capteur aux chocs, et parce qu'il était impossible de le fixer sur la baguette.

Finalement, seul un capteur magnétique d'angle et un accéléromètre dirigé selon l'axe tangentiel ont été embarqués sur la baguette. Ils sont suffisants, moyennant des traitements efficaces, pour capter l'ensemble des paramètres du mouvement jugés pertinents.

La figure IV.5 montre une des baguettes augmentées. Tous les capteurs étant fixés aux baguettes, le système a l'avantage d'être portable, et n'impose pas au percussionniste de frapper à un endroit donné.

IV.3.2 Interface

Nous disposons de deux interfaces de captation assez différentes pour réaliser les percussions augmentées. Si les algorithmes de traitement sont théoriquement indépendants de l'interface, leur implémentation pratique doit prendre en compte les spécificités de chacune d'elles. Deux caractéristiques principales rentrent en compte pour le choix de l'interface de captation :

– **Le format des données**

Pure-Data et Max/MSP sont capables de traiter deux types de données : des données de "contrôle" et des données audio. Les données de contrôle peuvent être traitées individuellement alors que les données audio sont regroupées en buffer.

Pour notre application, il est indispensable de travailler à l'échantillon près. Cela nous permet de garder une bonne précision temporelle, et de définir des filtres par leurs équations de récurrence.

Les données de contrôle sont directement accessibles à l'échantillon près, ce qui permet de leur appliquer très facilement des traitements de type filtrages récursifs. Au contraire, réaliser un traitement à l'échantillon près sur un buffer audio oblige à développer des programmes en C externes à Max/MSP.

Le Toaster renvoie l'information des capteurs sous la forme de données de contrôle, la Rimas Box sous la forme d'un flux audio, découpé en buffers. Du point de vue de l'implémentation des traitements, le Toaster est donc plus intéressant.

– **La fréquence d'échantillonnage**

Les deux interfaces ont une fréquence d'acquisition de données constante, contrairement aux interfaces événementielles de type MIDI. Cette propriété est fondamentale car elle rend possible le filtrage numérique de l'information des capteurs.

Le toaster échantillonne l'information des capteurs à 200 Hz. Aussi, on ne peut pas récupérer les données du geste qui dépassent 100 Hz. Cette fréquence est suffisante pour des données gestuelles classiques. Mais la percussion augmentée est un cas extrême, où le mouvement présente une discontinuité lorsque la baguette percute une surface rigide. Nous avons vu que le signal d'accélération tangentielle présentait alors des fréquences allant jusqu'à 1200 Hz.

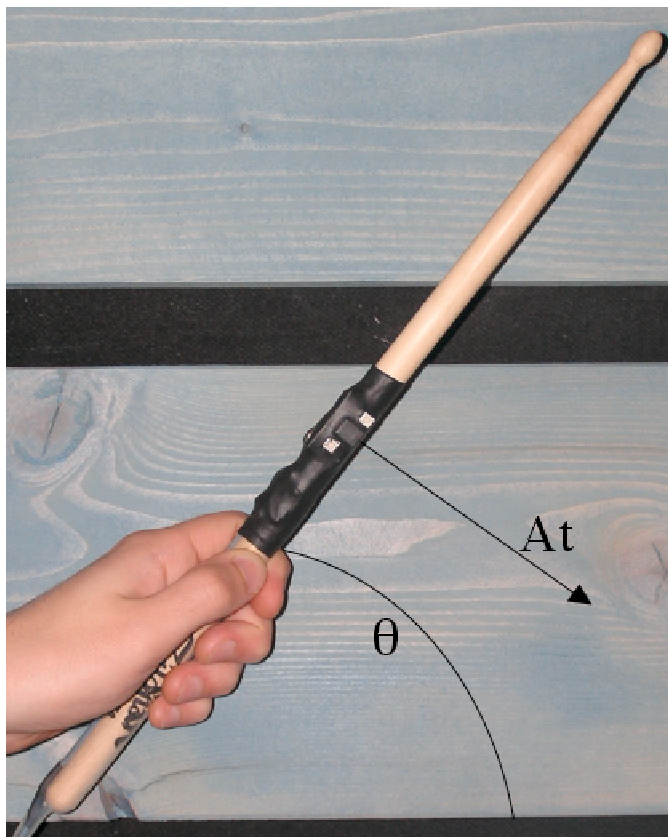
Cette particularité fréquentielle du signal au moment des chocs est très importante. Nous verrons au paragraphe VI.2 qu'elle permet de détecter les attaques du percussionniste de manière robuste. La percussion augmentée oblige donc à capter des hautes fréquences du signal, qui ne sont pas présentes dans un mouvement classique.

La Rimas Box échantillonnant les capteurs à 5512.5 Hz, elle permet d'obtenir tout le spectre de l'information gestuelle, même aux moments des impacts. Elle semble donc bien adaptée au cas particulier de la percussion augmentée.

Le format des données de la Rimas Box complique notablement l'implémentation des algorithmes de traitement du signal des capteurs. Cependant, sa fréquence d'échantillonnage élevée est un atout déterminant pour l'application des percussions augmentées. Nous avons travaillé avec les deux interfaces, mais la Rimas Box a permis d'aller plus loin pour la détection d'attaque.



(a) vue de dessus



(b) vue de profil avec les données captées

FIGURE IV.5: Une baguette augmentée

Chapitre V

Données de la Rimas Box et prétraitement

V.1 Format du signal de sortie de la Rimas Box

Comme nous l'avons mentionné auparavant, 8 capteurs en entrée de la Rimas Box se partagent un même convertisseur analogique/numérique fonctionnant à 44100 Hz. Chacun est donc échantillonné en réalité à 5512.5 Hz. Les problèmes de recouvrement spectral sont évités car la Rimas Box applique un filtre anti-repliement à 2000 Hz en amont sur les capteurs. Cependant, les signaux de sortie de la Rimas Box, récupérés dans Max/MSP, sont des flux audio à 44100 Hz. Ceux-ci contiennent bien l'information des capteurs, mais chaque point est répété 8 fois : le signal utile à 5512.5 Hz a été transformé en un signal en escalier à 44100 Hz (cf.figure V.1).

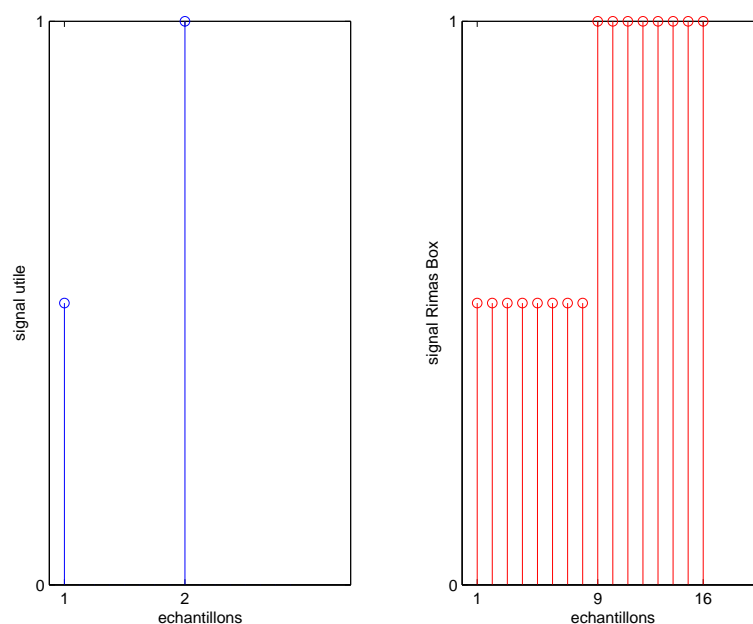
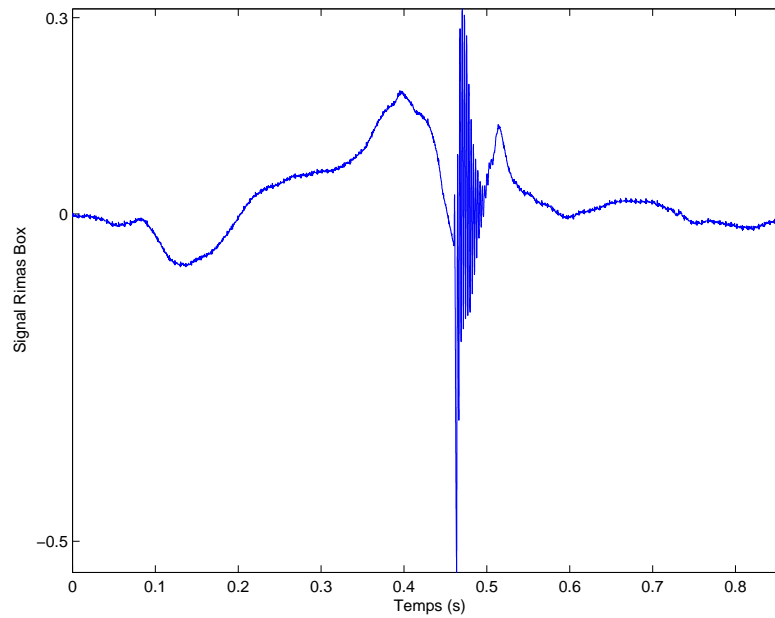


FIGURE V.1: Signal utile et signal de la Rimas Box

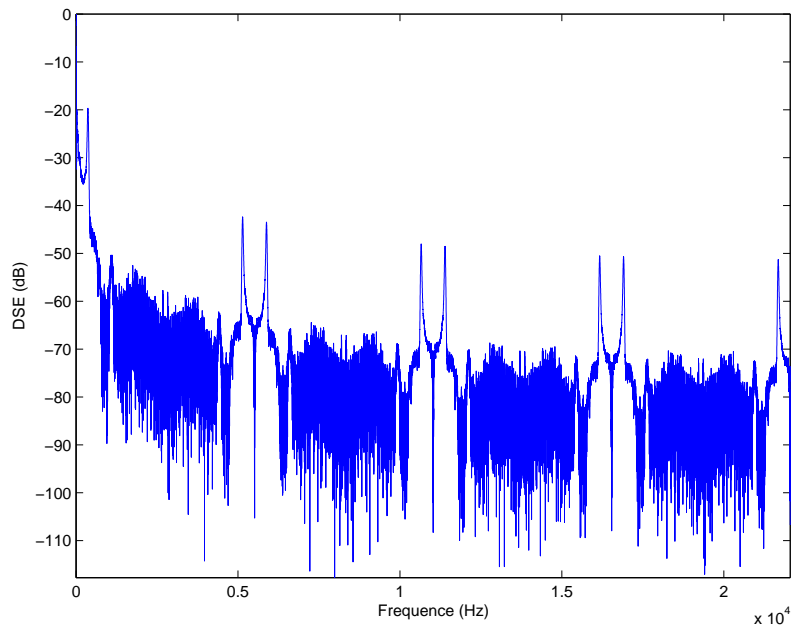
Examinons le phénomène en prenant un exemple de frappe du percussionniste. La figure V.2 montre le signal du capteur d'accélération tangentielle issu de la Rimas Box en fonction du temps.

FIGURE V.2: **Signal d'accélération issu de la Rimas Box**

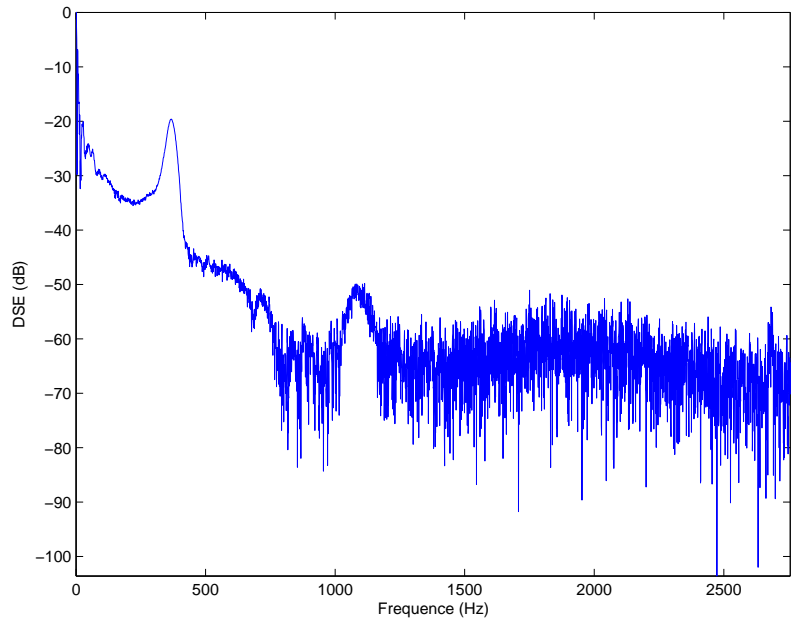
Compte tenu de la durée considérée (presque une seconde) il est impossible de déceler à l'oeil nu que chaque point est répété huit fois, mais un zoom sur la figure nous le montrerait (cf.figure V.1).

Plaçons nous du point de vue spectral : la figure V.3 montre en haut la Densité Spectrale d'Energie (DSE) du signal de la Rimas Box, et en bas, la DSE du même signal sous-échantillonné d'un facteur huit, c'est-à-dire celle du signal utile.

En comparant les deux DSE, on constate nettement que le signal de la Rimas Box contient des harmoniques supplémentaires. Un prétraitement est nécessaire pour récupérer uniquement le signal utile, avant de pouvoir lui appliquer d'autres transformations comme par exemple une dérivation.



(a) DSE du signal de la Rimas Box



(b) DSE du signal utile

FIGURE V.3: DSE du signal de la Rimas Box et du signal utile

V.2 Prétraitement : correction du signal de sortie de la Rimas Box

Plusieurs manipulations du flux audio sont possibles pour récupérer le signal utile. La première est de sous-échantillonner en temps réel le signal de la Rimas Box, la deuxième de le filtrer passe-bas.

V.2.1 Sous-échantillonnage

La méthode la plus évidente pour ne récupérer que les points pertinents du signal de la Rimas Box est de sous-échantillonner d'un facteur 8 en temps réel dans Max/MSP. Le logiciel permet d'effectuer ce type d'opération avec l'objet "poly", mais cela oblige à travailler dans des sous-patches, ce qui n'est pas toujours très pratique.

Nous avons également codé un objet externe pour Max/MSP qui permet de sous-échantillonner un flux audio tout en le transformant en un flux de messages de type flottant. Cet objet est très utile car il permet d'utiliser ensuite tous les traitements disponibles en Max pour les messages qui sont inutilisables avec des données de type signal. Cependant, Max/MSP est incapable de traiter les opérations sur les messages à des fréquences proches de l'audio. Pour que notre objet reste utilisable, il faut donc sous-échantillonner par un facteur élevé, ou disposer d'un processeur particulièrement puissant.

V.2.2 Filtrage passe bas

Une autre idée est de récupérer le spectre du signal utile par un filtrage passe-bas du flux audio de la Rimas Box, et de continuer à travailler à 44100 Hz. Nous décrivons ici le phénomène de manière théorique, pour montrer l'efficacité de cette méthode.

Soit x le signal utile échantillonné à 5512.5 Hz, y le même signal sur-échantillonné d'un facteur 8 et z le signal en sortie de la Rimas Box, échantillonné à 44100 Hz. Une représentation des signaux est donnée figure V.4.

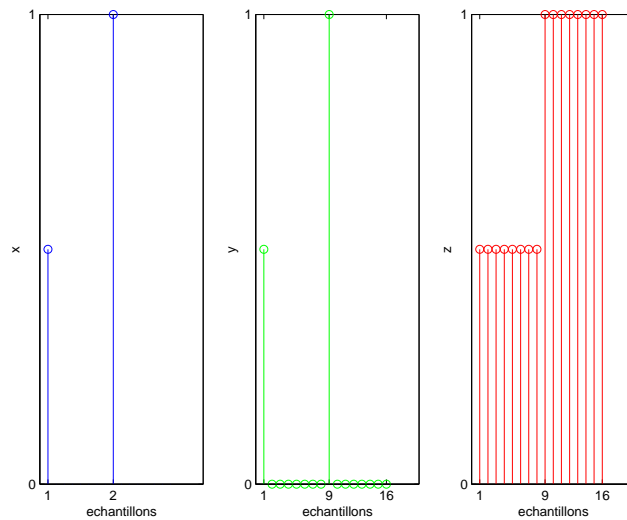


FIGURE V.4: Signal utile (x), sur-échantillonné (y) et signal de la Rimas Box (z)

Nous cherchons à exprimer le spectre de z en fonction de celui de x . Pour une démonstration plus formelle, nous désignerons le facteur de sur-échantillonnage par M , tout en sachant que pour notre application $M = 8$. Le signal sur-échantillonné y est défini par :

$$y : \begin{cases} y(M.n) = x(n) & \forall n \in \mathbb{Z} \\ y(n) = 0 & \text{ailleurs} \end{cases}$$

Calculons sa transformé en Z :

$$Y(z) = \sum_{k \in \mathbb{Z}} y(k).z^{-k}$$

$$\text{or} \quad y(k) = 0 \quad \forall k \notin M.\mathbb{Z}$$

$$\implies Y(z) = \sum_{k \in M.\mathbb{Z}} y(k).z^{-k}$$

$$\text{On pose } n = \frac{k}{M} : \quad Y(z) = \sum_{n \in \mathbb{Z}} y(M.n).z^{-M.n}$$

$$Y(z) = \sum_{n \in \mathbb{Z}} x(n).(z^M)^{-n}$$

$$Y(z) = X(z^M)$$

Ce qui donne pour les Transformées de Fourier à Temps Discret (TFTD) des signaux x et y :

$$Y(e^{2j\pi\nu}) = X(e^{2j\pi M\nu}) \quad (\text{V.1})$$

La TFTD du signal sur-échantillonné y est donc celle du signal utile x compressée par le facteur de sur-échantillonnage en fréquence réduite. Si Fe est la fréquence d'échantillonnage du signal x et $M.Fe$ celle de y , le spectre de y est alors celui de x périodisé tous les Fe en fréquence réelle.

Illustrons ce résultat en reprenant l'exemple du signal d'accélération utilisé au paragraphe V.1. La figure V.5 donne la DSE du signal utile, celui-ci étant échantillonné à 5512.5 Hz. La figure V.6 donne la DSE en fréquence réelle du même signal sur-échantillonné d'un facteur 8. La nouvelle fréquence d'échantillonnage est donc considérée comme étant 44100 Hz. On constate qu'il s'agit bien du même spectre périodisé tous les 5512.5 Hz.

Continuons pour caractériser le spectre du signal de la Rimas Box z . On considère la fenêtre rectangulaire w de taille M :

$$w(n) = \sum_{k=0}^{M-1} \delta(n-k) \quad \text{où } \delta \text{ est le symbole de Kronecker}$$

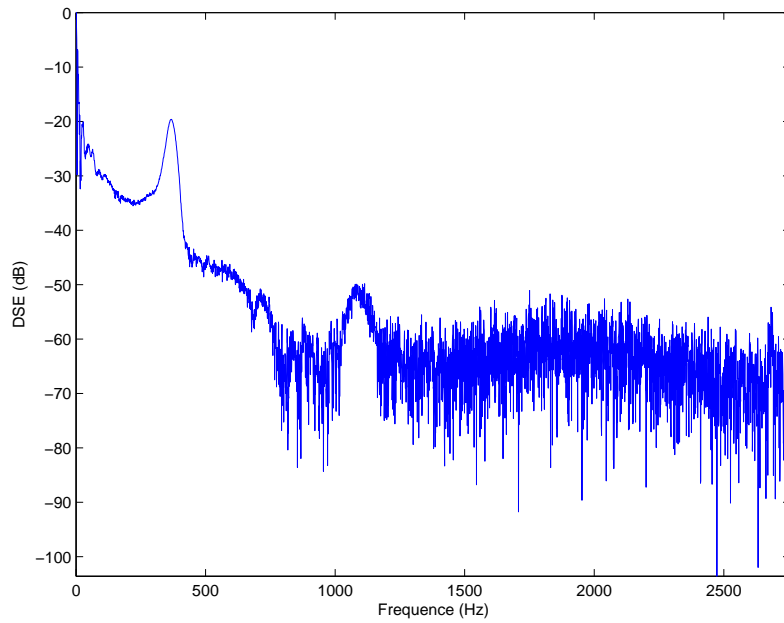


FIGURE V.5: DSE du signal utile

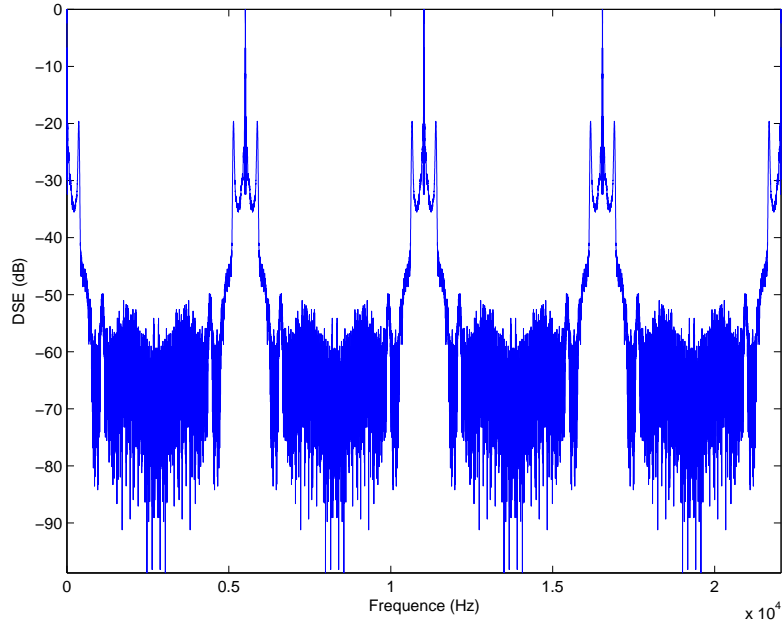


FIGURE V.6: DSE du signal sur-échantillonné

La TFTD de la fenêtre rectangulaire w vaut :

$$W(e^{2j\pi\nu}) = e^{-j\pi(M-1)\nu} \cdot \frac{\sin(M\pi\nu)}{\sin(\pi\nu)} \quad (\text{V.2})$$

Or :

$$\begin{aligned} z(n) &= \sum_{k=0}^{M-1} y(n) \star \delta(n-k) = y \star w(n) \\ \Rightarrow Z(e^{2j\pi\nu}) &= Y(e^{2j\pi\nu}) \times W(e^{2j\pi\nu}) \end{aligned} \quad (\text{V.3})$$

Ainsi, en injectant les formules V.1 et V.2 dans V.3, on obtient :

$$Z(e^{2j\pi\nu}) = X(e^{2j\pi M\nu}) \cdot e^{-j\pi(M-1)\nu} \cdot \frac{\sin(M\pi\nu)}{\sin(\pi\nu)}$$

La TFTD du signal de la Rimas Box z est donc celle du signal utile x , mais compressée d'un facteur 8 en fréquence réduite et pondérée par la TFTD de la fenêtre rectangulaire w . Revenons à l'exemple du signal d'accélération étudié précédemment. La figure V.7 montre la DSE du signal tel qu'il est récupéré à la sortie de la Rimas Box.

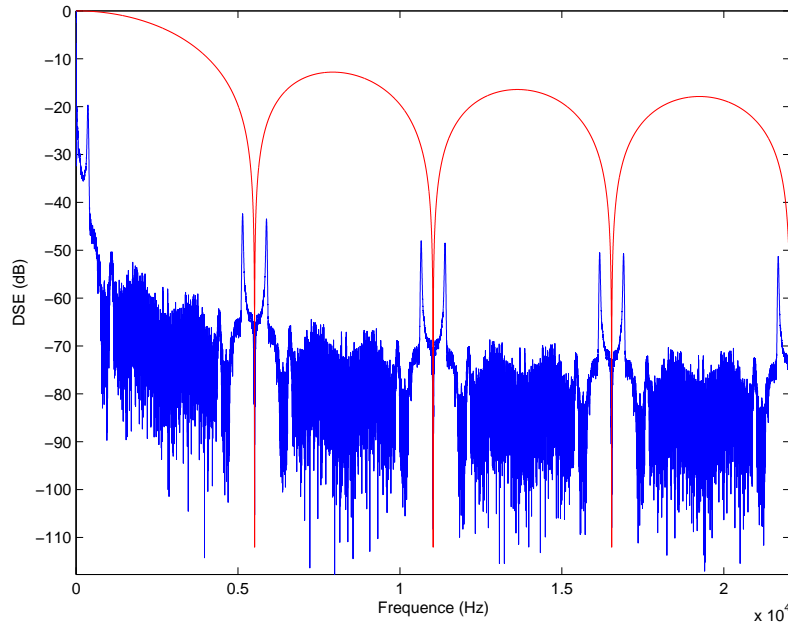


FIGURE V.7: DSE du signal de la Rimas Box et de la fenêtre

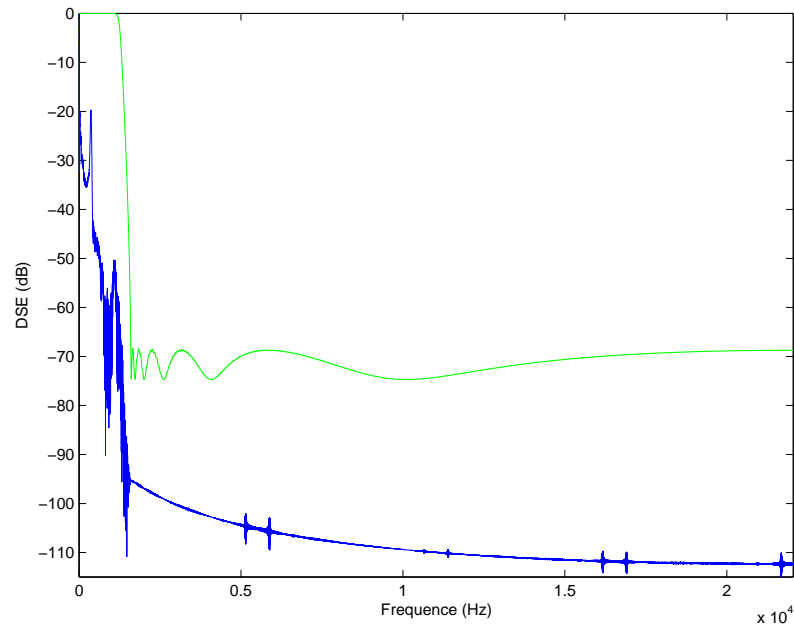
La DSE est bien la même que celle du signal sur-échantillonné (cf. figure V.6), mais pondéré par le spectre de la fenêtre rectangulaire, représenté en rouge sur la figure V.7.

On constate par ailleurs que le spectre de la fenêtre w est plat en dessous de 2000 Hz, domaine de fréquence du signal utile. Par conséquent, on peut récupérer tout le spectre du signal utile en appliquant au signal de la Rimas Box un filtrage passe-bas qui coupe à 2000 Hz.

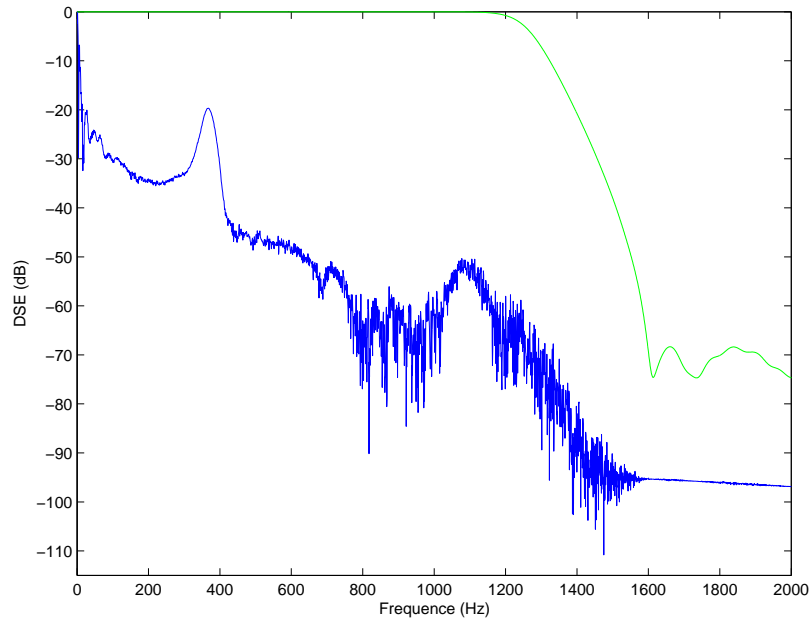
Toute l'énergie du signal utile étant concentrée en dessous de 1200 Hz, nous appliquons plutôt un filtre qui coupe tout à partir de cette fréquence, ce qui permet d'éliminer le bruit en hautes fréquences.

La figure V.8 donne la fonction de transfert du filtre et le spectre du signal de la Rimas Box après filtrage. Elle montre que l'on retrouve bien le spectre du signal utile en comparant avec la figure V.5.

La figure V.9 donne en fonction du temps le signal utile, le signal en sortie de la Rimas Box et le signal de la Rimas Box filtré. La forme d'onde apparaît bien conservée après le filtrage. Pour constater visuellement une différence, il faudrait observer les signaux à l'échantillon près. On peut cependant constater sur le zoom un léger décalage du signal filtré, de l'ordre de la milliseconde, dû au retard de groupe du filtre passe-bas.



(a) fréquences de 0 à 22500 Hz



(b) zoom sur les fréquences < 2000 Hz

FIGURE V.8: DSE du signal de la Rimas Box filtré et fonction de transfert du filtre

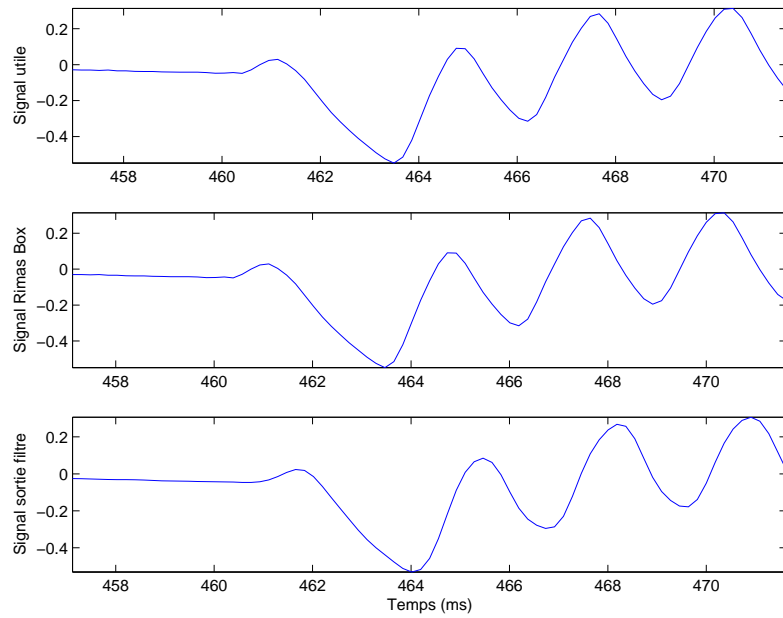
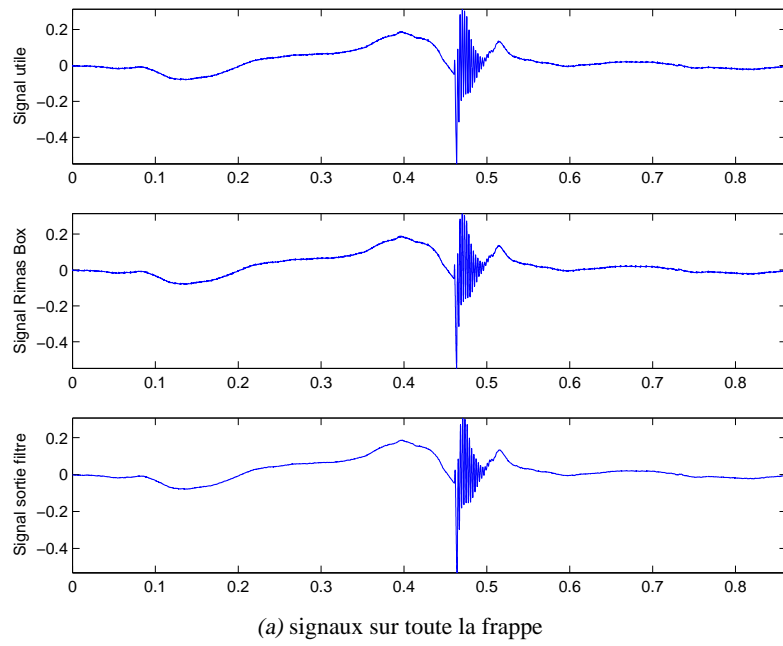


FIGURE V.9: Signal de la Rimas Box, signal utile et signal de la Rimas Box filtré

Chapitre VI

Traitements des données gestuelles

L'information des capteurs n'est pas utilisable directement pour la synthèse sonore. Nous avons déjà évoqué les pré-traitements imposés au signal en sortie de la Rimas Box. Nous allons maintenant décrire les opérations qui permettent d'extraire les paramètres du geste qui nous intéressent : la vitesse angulaire des baguettes, la détection des impacts et de leur intensité. Pour ce qui est de la détermination de l'angle entre la baguette et l'horizontal, le simple filtrage passe-bas effectué en pré-traitement sur le capteur d'angle est suffisant pour obtenir une information exploitable. Nous n'en reparlerons donc pas ici.

VI.1 Détermination de la vitesse angulaire des baguettes

Nous avons expliqué au paragraphe IV.3.1 que la technique la plus appropriée pour obtenir la vitesse angulaire d'une baguette est de dériver l'information du capteur d'angle. L'opération de dérivation étant très sensible au bruit, nous appliquons auparavant un filtrage passe-bas sur le capteur, qui coupe à 20 Hz. Le filtre passe-bas a été obtenu par la méthode de Chebyshev à l'aide de la toolbox "Filter Design" de Matlab. Il s'agit d'un filtre ARMA d'ordre 7. Nous avons implémenté sa décomposition en biquads dans Max/MSP pour filtrer les données en temps réel (cf. figure VI.1). La figure VI.2 donne la réponse en fréquence du filtre.

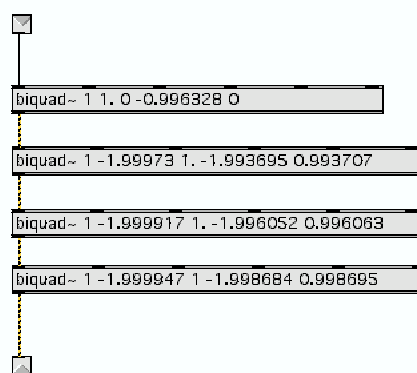


FIGURE VI.1: Implémentation en biquads dans Max/MSP

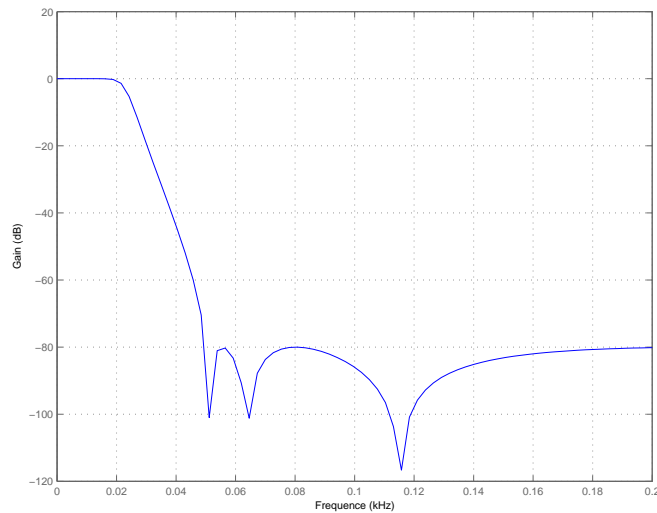


FIGURE VI.2: Filtre passe-bas

Le signal d'angle étant très basse fréquence, on ne perd pas d'information. Le bruit est par contre sensiblement réduit. La figure VI.3 montre les opérations de filtrage et de dérivation sur un exemple de frappe.

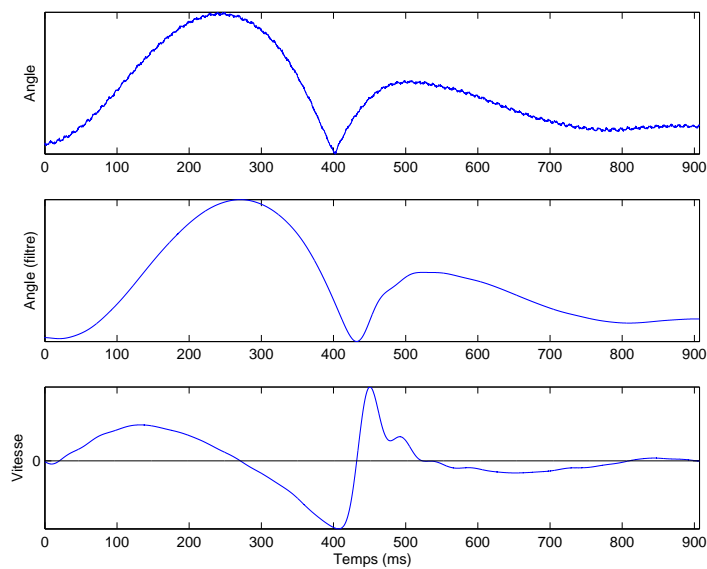


FIGURE VI.3: Filtrage du capteur d'angle et obtention de la vitesse

On constate un retard d'une trentaine de millisecondes imposé par le filtrage à 20 Hz. Ce délai est sans conséquence car il concerne seulement l'avant et l'après-geste de l'instrumentiste, qui s'effectuent souvent sur des durées supérieures à la seconde. De plus ces types de gestes ne permettent pas de prendre des repères perceptifs précis dans le temps, contrairement à l'attaque. La latence est donc transparente pour le percussionniste.

Cette méthode donne une valeur relative de la vitesse angulaire des baguettes, qui permet une segmentation intéressante du jeu du percussionniste. On peut séparer les phases du mouvement où la baguette descend de celles où elle monte. De plus, la valeur de la vitesse peut être utilisée comme paramètre de synthèse sonore, afin d'obtenir un son avant même que la baguette ait frappé la percussion.

VI.2 Détection de l'attaque

La détection des impacts des baguettes sur une surface rigide (détection d'attaque) est un problème un peu plus compliqué que de déterminer leurs vitesses angulaires. Nous avons cherché à isoler des paramètres robustes significatifs des chocs. Les mesures décrites au paragraphe IV.2 ont montré que l'attaque se ressent particulièrement sur l'accélération tangentielle des baguettes, ce qui se confirme au niveau théorique. Nous reprenons pour cela les formules et la figure du paragraphe IV.1.1 :

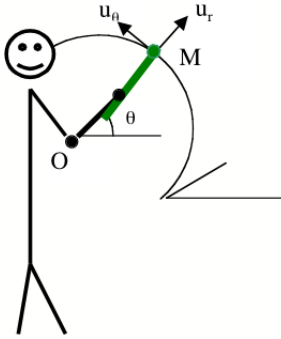


FIGURE VI.4: **Modèle unidimensionnel du mouvement de la baguette**

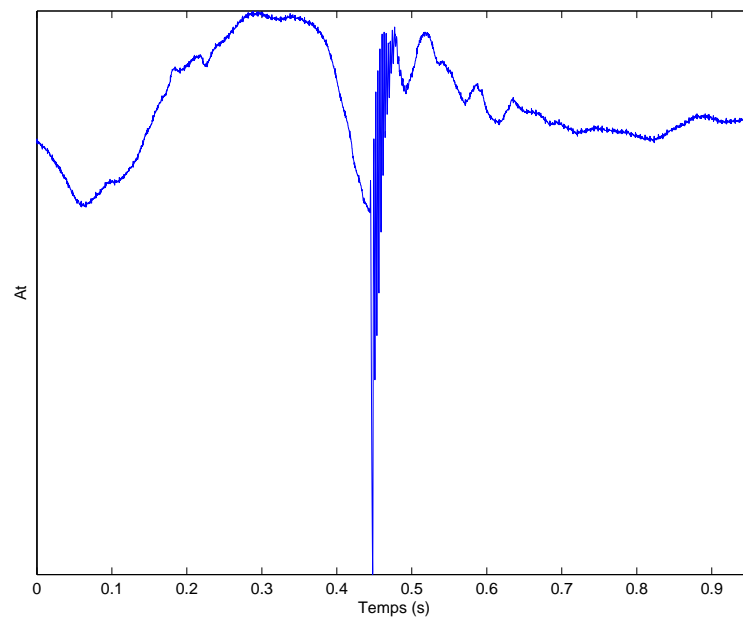
$$\overrightarrow{OM} = r\vec{u}_r \quad (\text{VI.1})$$

$$\vec{V} = \frac{d\overrightarrow{OM}}{dt} = r\dot{\theta}\vec{u}_\theta \quad (\text{VI.2})$$

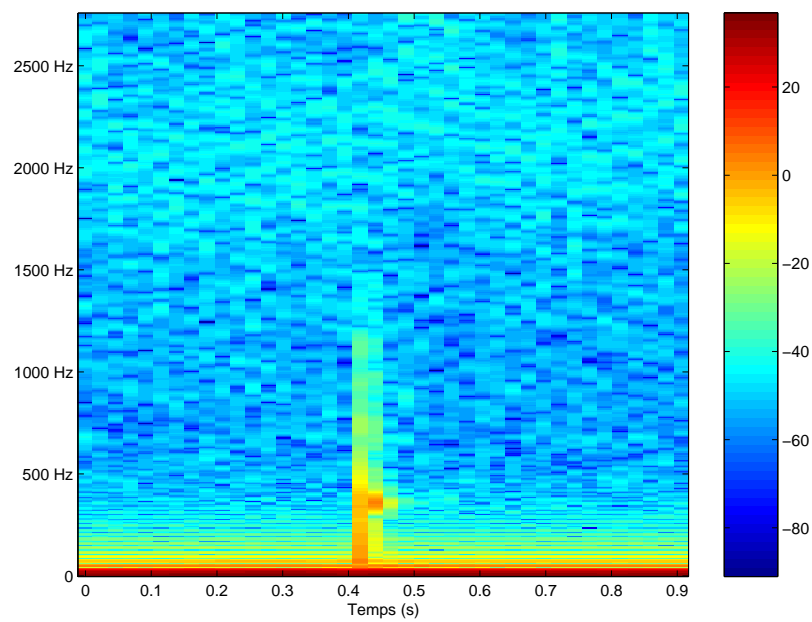
$$\vec{A} = \frac{d\vec{V}}{dt} = r\dot{\theta}^2\vec{u}_r + r\ddot{\theta}\vec{u}_\theta \quad (\text{VI.3})$$

Considérons un modèle simpliste de l'impact où la vitesse angulaire $\dot{\theta}$ subit une discontinuité dès lors que la baguette rencontre une surface rigide. Autrement dit, la vitesse prend la forme d'une fonction échelon. On obtient alors une accélération tangentielle en $\ddot{\theta}$ égale à un dirac aux moments des chocs, signal d'amplitude infinie et excitant toutes les fréquences.

Dans la réalité, la vitesse n'est pas discontinue mais présente obligatoirement une forte variation aux moments des impacts. L'accélération tangentielle présente elle bien des pics notables et des hautes fréquences aux instants des chocs, dans les mesures effectuées, comme le montre la figure VI.5.

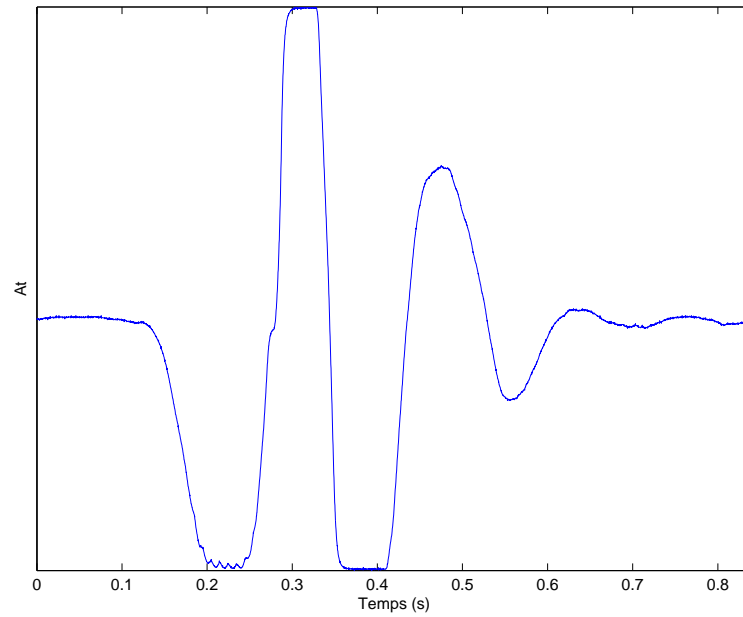


(a) signal en fonction du temps

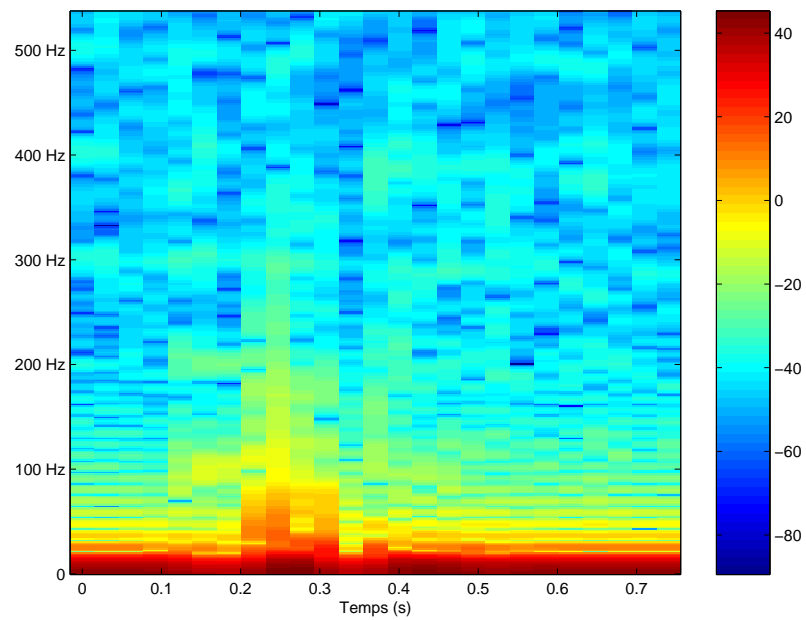


(b) spectrogramme

FIGURE VI.5: Accélération tangentielle sur une frappe avec impact réel



(a) signal en fonction du temps



(b) spectrogramme

FIGURE VI.6: Accélération tangentielle sur une frappe dans le vide

La figure VI.5 montre les données d'accélération tangentielle en fonction du temps pour une frappe sur une surface réelle, et le spectrogramme associé. La fenêtre utilisée pour la transformée de Fourier à court terme est une fenêtre de Hanning de 128 points. Chaque transformée de Fourier discrète (TFD) est calculée sur 1024 points fréquentiels et le recouvrement temporel entre deux fenêtres successives est de 64 points.

On peut voir que le signal d'accélération tangentielle présente de l'énergie dans des fréquences supérieures à 1000 Hz aux moments des chocs.

Notre système doit détecter uniquement les impacts sur une surface réelle, et ignorer les frappes dans le vide. Examinons le signal d'accélération tangentielle lorsque le percussionniste frappe dans le vide avec la figure VI.6. Nous avons limité la représentation du spectrogramme aux fréquences inférieures à 500 Hz, car il n'y a pas d'énergie dans les fréquences plus élevées. Nous avons utilisé une fenêtre de Hanning de 512 points pour avoir une précision fréquentielle correcte.

On constate que s'il n'y a plus d'impact réel, le signal présente tout de même des fréquences plus élevées aux moments où la vitesse change de sens, mais celles-ci ne dépassent pas 200 Hz.

La première idée pour isoler les impacts consiste simplement à poser un seuil sur les données temporelles brutes du capteur d'accélération tangentielle. Dès que le signal dépasse ce seuil on considère qu'il y a choc. Cependant cette technique n'est absolument pas robuste. Tout d'abord l'offset du capteur d'accélération est variable, ce qui pose problème pour fixer le seuil. D'autre part, l'accélération peut atteindre une forte amplitude même lorsqu'il n'y a pas de choc réel, et déclencher de fausses détections.

Une meilleure idée pour détecter l'attaque consiste à tirer partie des particularités fréquentielles du signal aux instants des impacts. L'examen des figures VI.5 et VI.6 nous montre que le signal présente de l'énergie au-delà de 200 Hz, uniquement au moment des impacts réels. Nous avons conçu notre algorithme de détection sur cette considération.

L'algorithme de détection d'attaque : par filtrage, on recueille en temps réel la puissance du signal dans les fréquences supérieures à 200 Hz. Si elle dépasse un seuil, on considère qu'il y a choc.

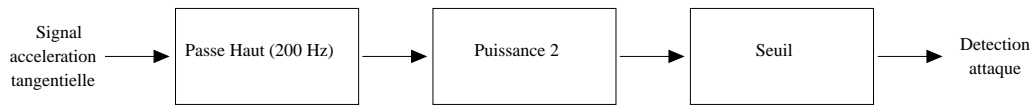


FIGURE VI.7: Algorithme de détection d'attaque

L'utilisateur peut ajuster le seuil de détection pour obtenir une détection plus ou moins sensible. La puissance au-dessus de 200 Hz est obtenue en passant le signal dans un filtre passe-haut, qui coupe les fréquences inférieures à 200 Hz, et en élevant le signal de sortie au carré.

La figure VI.8 donne la réponse en fréquence du filtre passe-haut.

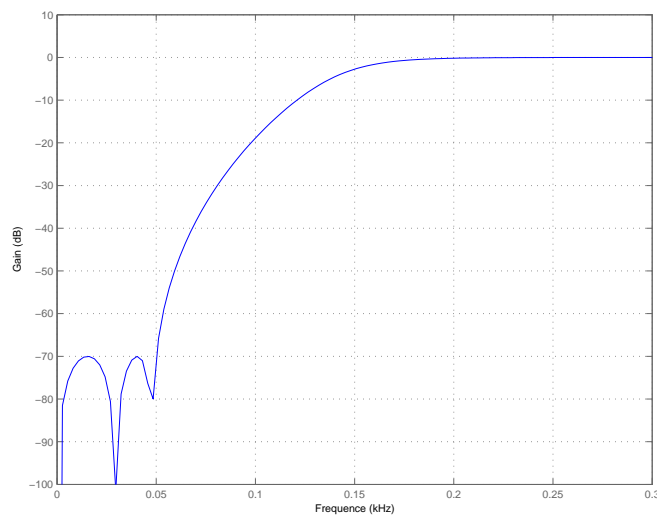


FIGURE VI.8: Filtre passe-haut

Ce filtre a encore été obtenu par la méthode de Chebyshev à l'aide de la toolbox "Filter Design" de Matlab. Il s'agit d'un filtre ARMA d'ordre 5. Nous avons implémenté sa décomposition en biquads dans Max/MSP pour filtrer les données en temps réel (cf. figure VI.9).

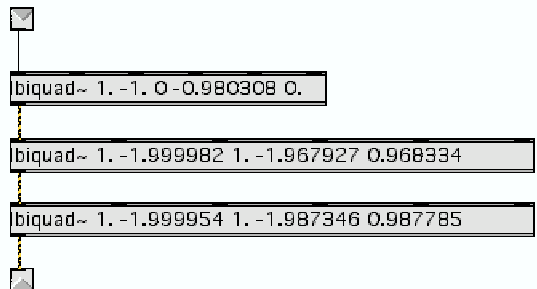
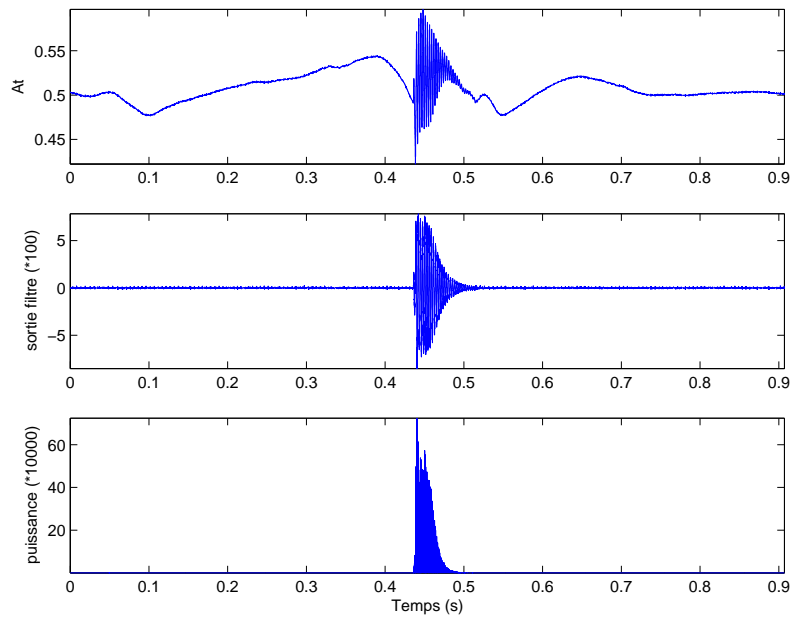
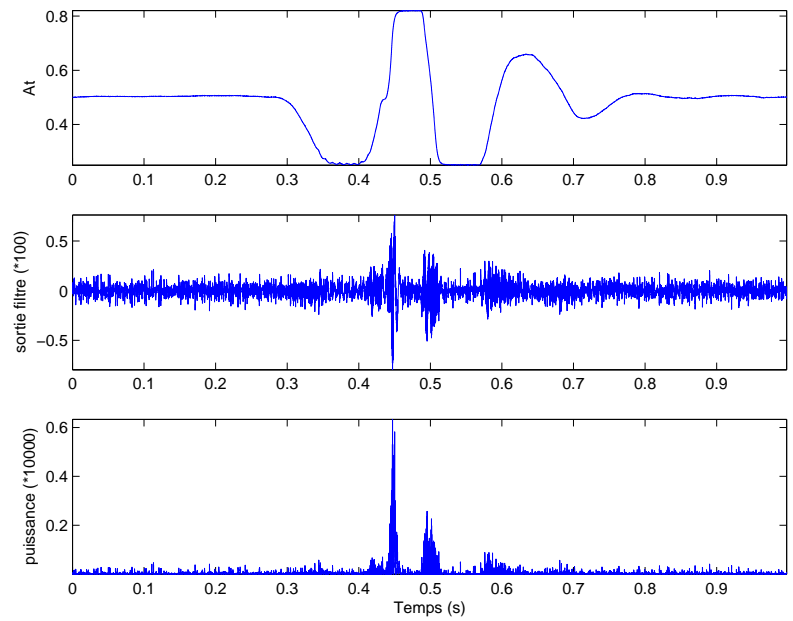


FIGURE VI.9: Implémentation en biquads dans Max/MSP



(a) frappe avec impact réel



(b) frappe dans le vide

FIGURE VI.10: Accélération tangentielle, résultat du filtrage PH et puissance du signal filtré

La figure VI.10 montre en haut l'accélération tangentielle sur une frappe avec impact réel, le résultat du filtrage passe-haut, ainsi que la puissance instantanée du signal filtré. On trouve en bas les mêmes données mais captées cette fois lors d'une frappe dans le vide, sans impact réel. En examinant cette figure, on voit que la puissance en sortie du filtre passe-haut atteint des valeurs élevées uniquement à l'instant du choc réel. Mettre un seuil sur cette puissance pour détecter les impacts est donc une technique fiable.

La figure VI.11 montre un zoom sur la frappe réelle de la figure VI.10 au moment de l'impact. Le seuil de détection de choc appliqué sur la puissance apparaît en rouge. Cette valeur du seuil est donnée à titre d'exemple. En réalité elle peut être réglée pour s'adapter à la sensibilité du percussionniste.

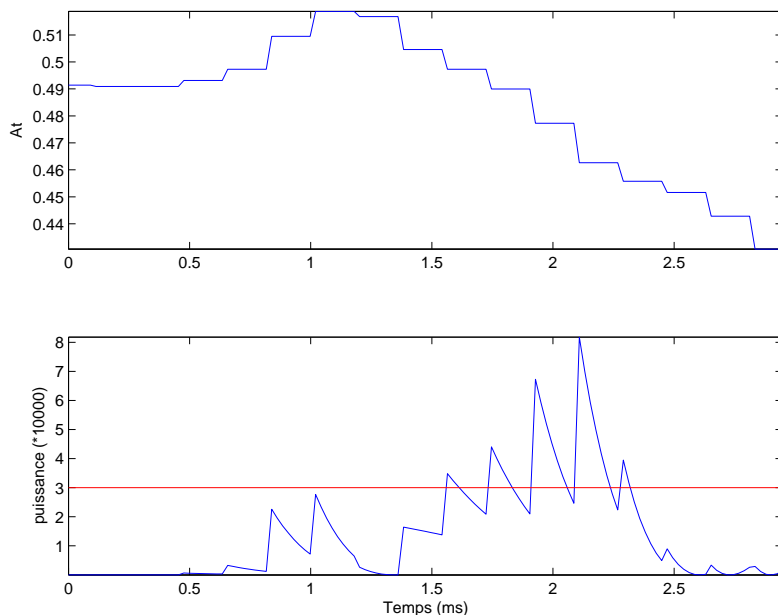


FIGURE VI.11: Accélération tangentielle sur un impact réel et seuil de détection sur la puissance

On voit sur cet exemple que la latence du système est de l'ordre de la milliseconde. L'impact correspond à la décroissance forte de l'accélération tangentielle. La détection du choc est réalisée lorsque la puissance de sortie du filtre dépasse le seuil, soit dès le début de cette décroissance.

L'algorithme s'avère très robuste en pratique. Le bruit en sortie du filtre est très faible, ce qui permet de détecter des chocs très petits.

Les fausses détections sont très rares, car il faut qu'il y ait réellement un choc pour exciter des fréquences supérieures à 200 Hz, et récupérer une puissance non négligeable en sortie du filtre. Il peut néanmoins arriver que la baguette heurte la paume de l'instrumentiste et déclenche un son pendant l'avant ou l'après-geste.

La sensibilité est réglable par le biais du seuil, ce qui permet de capter les nuances les plus subtiles d'un percussionniste, tout en se protégeant des fausses détections.

VI.3 Détermination de la vitesse des coups

Nous sommes maintenant en mesure de détecter en temps réel les impacts des baguettes sur une surface solide. Cette détection peut servir à déclencher des lectures d'événements sonores synchronisées sur le son acoustique d'une percussion réelle. Cependant, pour établir une superposition pertinente du son naturel de l'instrument et du son synthétique, il faut corréler leurs niveaux sonores. Autrement dit, la synthèse doit épouser les nuances de jeu du percussionniste, qui représentent les différents degrés d'intensité par lesquels peut passer le son. En informatique musicale, on emploie le terme de vitesse.

Malheureusement, le lien entre la puissance sonore dégagée et la force de frappe varie avec la surface percutée. Une membrane plus ou moins tendue, une symbale métallique ou une clave de bois auront des réactions différentes. Nous avons donc cherché à obtenir le meilleur compromis possible, qui soit cohérent au niveau perceptif pour le percussionniste. Celui-ci doit simplement "sentir" que le volume de la synthèse sonore s'accorde bien avec la force de sa frappe.

Rappelons que le capteur de pression a été jugé inadéquat pour notre application, pour des raisons de solidité, de confort pour l'instrumentiste (cela l'oblige à frapper à un endroit donné, puisqu'il est difficile de fixer le capteur au bout de la baguette) et surtout de temps de réponse. Il faut donc déterminer la vitesse des coups à partir des seuls capteurs d'accélération et d'angle intégrés à la baguette.

Nous avons mené différents essais pour obtenir un volume cohérent perceptivement à partir de ces données. La solution la plus intéressante est d'utiliser la vitesse angulaire de la baguette juste avant l'impact. Plus précisément, le volume est la vitesse angulaire de la baguette 30 ms avant le choc, élevée à la puissance $4/3$.

Nous avons choisi ce laps de temps essentiellement pour simplifier l'implémentation dans Max/MSP. Nous avons évoqué au paragraphe VI.1 le retard d'une trentaine de millisecondes du signal de vitesse dû au filtrage passe-bas utilisé. Le volume est donc obtenu simplement en allant lire la valeur courante de la vitesse à chaque détection d'attaque, et en l'élevant à la puissance $4/3$.

Nous avons également pensé à introduire une table pour commander le volume. L'instrumentiste peut éditer la table, pour adapter le volume des percussions augmentées à sa sensibilité.

VI.4 Réalisation en Max/MSP

Nous donnons ici simplement un aperçu des patches implémentés en Max/MSP pour réaliser les traitements décrits plus haut.

La figure VI.12 montre le module de traitement qui permet d'obtenir l'angle, la vitesse angulaire, et qui gère également la détection d'attaque.

La figure VI.13 montre un patch plus haut niveau, qui capture les données d'entrées, les envoie dans le module de traitement, passe ensuite les données traitées dans des modules de synthèse sonore et envoie le résultat à la carte son de l'ordinateur.

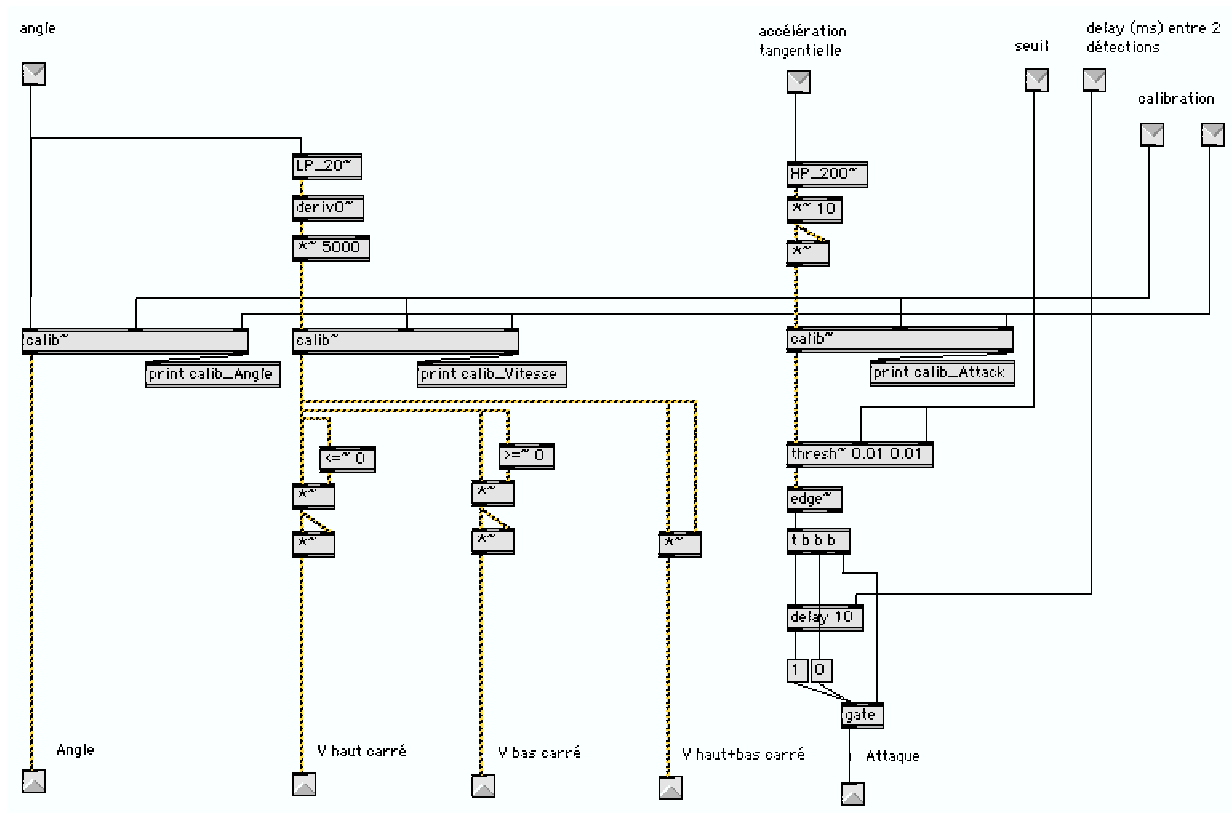


FIGURE VI.12: Module de traitement en Max/MSP

Chapitre VII

Mapping des données pour une synthèse sonore

Si la détection d'attaque permet le déclenchement d'événements musicaux, les données concernant l'ensemble du geste du percussionniste peuvent être utilisées pour contrôler les paramètres d'une synthèse sonore. Toute la question est alors de savoir quelle donnée va contrôler quel paramètre, pour obtenir un résultat intéressant du point de vue artistique. Une correspondance, ou mapping, doit être définie entre les informations gestuelles et les outils de synthèse.

VII.1 Définition

M. Wanderley définit le mapping dans [14] comme une stratégie de correspondance entre les variables de sortie d'un système et les entrées d'un dispositif de synthèse sonore. La figure VII.1 illustre cette définition.

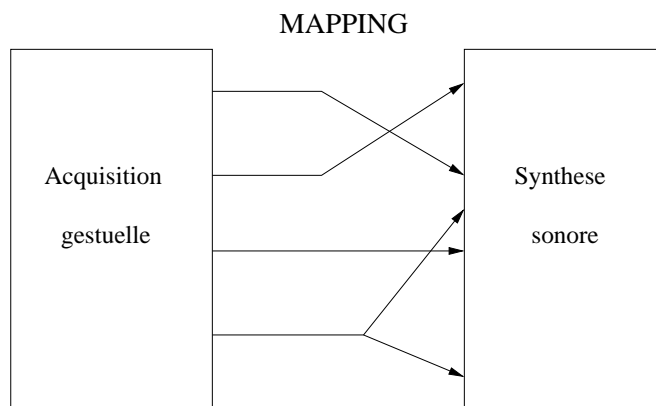


FIGURE VII.1: **Mapping**

Wanderley dégage deux points de vue sur le rôle du mapping dans le contexte des instruments augmentés ou virtuels.

Selon le premier point de vue, le mapping apparaît comme un choix esthétique du compositeur pour une oeuvre musicale. Ce dernier détermine la correspondance entre les actions de l'instrumentiste et le résultat sonore pour conduire l'oeuvre selon son goût artistique. Le mapping est alors un processus compositionnel.

Le second point de vue consiste à considérer le mapping comme une partie intégrante de l'instrument, indépendamment des oeuvres dans lesquelles il peut être utilisé. L'instrumentiste et le compositeur doivent s'adapter à ses possibilités.

On peut aussi distinguer deux principaux types de mapping utilisés dans la captation du geste :

- L'utilisation de réseaux de neurones et la reconnaissance de mouvement.
- L'utilisation d'une stratégie explicite de mapping.

Le premier type fait intervenir une adaptation interne du système au moyen d'un apprentissage (cf. [7]). Le second relie directement les données gestuelles aux paramètres de synthèse sonore.

VII.2 Application pour la percussion augmentée

Pour notre application des percussions augmentées, les données de sortie du système de captation sont :

- L'angle entre la baguette et le plan horizontal.
- La vitesse angulaire de la baguette.
- La détection d'attaque.
- L'intensité des impacts.

Nous avons imaginé plusieurs possibilités de mapping explicite, qui relient les données gestuelles à des paramètres de synthèse après une simple mise à l'échelle.

On peut par exemple contrôler la fréquence modulante et l'index d'une modulation de fréquence, les coefficients d'un filtre pour une synthèse source-filtre, la taille des grains d'une synthèse granulaire... Les données gestuelles peuvent également piloter des effets, comme un temps de réverbération, des paramètres de spatialisation du son, la vitesse de lecture d'un échantillon sonore... Bien évidemment, les possibilités sont infinies. Et des logiciels ouverts comme Pure-Data ou Max/MSP permettent d'implémenter très facilement de telles stratégies de mapping explicites.

Nous présentons ici quelques exemples qui ont été réalisés pendant le stage :

- La détection d'attaque permet le déclenchement d'échantillons sonores.
- La vitesse angulaire contrôle la fréquence et l'amplitude de sinusoides pour une synthèse additive.
- La vitesse angulaire contrôle le facteur de qualité d'un filtre dans lequel on passe un bruit blanc.
- L'angle contrôle la fréquence de coupure du filtre évoqué précédemment.

Ces exemples de mapping sur l'avant-geste sont cependant plus intéressants d'un point de vue démonstratif qu'artistique. Par ailleurs, nous n'avons pas eu le temps de nous pencher sur un mapping plus complexe, qui permettrait une reconnaissance de frappe grâce à un réseau de neurone par exemple. Ce sujet est évoqué dans les perspectives d'évolution du système au chapitre VIII.1.

Soulignons simplement que la percussion augmentée apporte au compositeur un champ de possibilités immense. Pouvoir déclencher un son avant la frappe du percussionniste donne de nouvelles perspectives esthétiques à explorer dans le cadre d'une oeuvre musicale.

Chapitre VIII

Perspectives

La création d'instruments augmentés soulève des problématiques complexes. Le temps imparti pour le stage n'a pas suffi à la réalisation de certaines pistes qui mériteraient d'être explorées, et que nous évoquons ici.

VIII.1 Classification et reconnaissance des frappes

Pour notre étude, nous avons préféré dans un premier temps laisser de côté la classification des frappes. Daniel Champolini préférerait une synthèse sonore utilisant de manière explicite les données gestuelles comme la vitesse ou l'orientation de la baguette.

Cependant, une autre stratégie de mapping pourrait être explorée, mettant en jeu des outils comme les réseaux de neurones pour une reconnaissance des frappes. Cette idée se rapproche des recherches actuelles de La kitchen, décrites dans l'article [7]

Evoquons aussi les travaux menés au Scrim par Myriam Desainte-Catherine. On peut trouver dans [9] une classification des frappes en trois types :

- Up : la baguette part d'une position haute, descend puis remonte pour arriver à une position basse.
- Down : la baguette part d'une position basse, descend puis remonte pour arriver à une position haute.
- Piston : la baguette descend puis remonte à la position de départ.

Ce modèle a été réalisé pour des frappes dans le vide, et ne correspond pas nécessairement à une situation de jeu sur une percussion réelle. On peut cependant imaginer de l'adapter pour notre percussion augmentée.

VIII.2 Captation d'autres paramètres du mouvement

Notre système ne capte actuellement qu'une dimension du mouvement de la baguette avec l'angle θ . Il pourrait être intéressant de capter d'autres paramètres, pour une analyse plus fine des gestes du percussionniste.

On peut penser par exemple à poser un second capteur d'angle sur le poignet du percussionniste. La différence entre les deux angles mesurés peut apporter une nouvelle information intéressante sur le geste instrumental.

Ce type de données permettrait d'obtenir une catégorisation plus complète des différentes frappes effectuées en situation de jeu réel. On peut également les utiliser pour étendre les possibilités de l'instrument augmenté, et le faire réagir à des gestes particuliers, qui ne relèvent pas d'une situation de jeu classique. Un nouveau langage gestuel peut ainsi être créé, en même temps qu'un nouvel instrument.

Enfin, une piste intéressante serait d'ajouter un système pour détecter la position de la baguette dans un repère absolu. Ainsi, les sons de synthèse déclenchés par les impacts des baguettes pourraient varier en fonction de l'endroit percuté. Un outil comme le "Flock of bird" ou une capture video, couplés à notre percussion augmentée, la feraient évoluer vers une percussion virtuelle complète.

Conclusion

Le but du stage était la conception d'un système de percussions augmentées à l'aide d'outils de captation gestuelle. A l'issue des quatre mois, nous sommes arrivés à la réalisation d'un prototype convaincant. Des capteurs d'angle et d'accélération fixés sur deux baguettes de percussion captent les gestes de l'instrumentiste. Leur information est numérisée et transmise à un ordinateur grâce à une interface sensible. Ces données pilotent en temps réel une synthèse sonore.

Ce matériel de captation gestuel a le grand avantage de proposer des fréquences d'échantillonnage de plusieurs centaines de Hertz, nécessaires pour l'application des percussions augmentées compte tenu des fréquences spécifiques mises en jeu.

Le système réalisé est simple et confortable pour l'instrumentiste puisque tous les capteurs sont embarqués sur les baguettes. Il permet de détecter les impacts des baguettes sur un objet solide quelconque, pour piloter le déclenchement d'événements sonores. Cet aspect constitue un contrôleur gestuel binaire d'une remarquable précision temporelle. Un traitement supplémentaire permet de mesurer la vélocité du coup porté, pour contrôler le volume de manière cohérente avec la force de frappe. En situation de jeu sur une percussion réelle, chaque coup porté par le percussionniste peut déclencher un son synthétique, parfaitement synchronisé avec le son naturel de l'instrument. On obtient une percussion au son uniquement synthétique en tapant sur une surface produisant peu de son naturellement, comme un bloc de mousse assez rigide.

L'autre intérêt majeur du système est qu'il permet de capter le mouvement de l'instrumentiste dans sa globalité, pendant les phases d'avant et d'après-geste, ce qui va beaucoup plus loin que la simple détection d'attaque. Le déplacement et la vitesse angulaires des baguettes sont connus à tout instant, et pilotent une synthèse sonore. Ainsi on peut générer du son avant et après l'impact des baguettes sur la percussion.

Cet aspect offre au percussionniste et au compositeur des champs d'expression inédits. Une stratégie de mapping riche sur le plan artistique doit maintenant être développée. Un nouveau langage gestuel peut naître, sortant du cadre instrumental classique, pour tirer profit des possibilités offertes par les percussions augmentées.

Bibliographie

- [1] R. Avizienis and al. Scalable connectivity processor for computer music performance systems. In *International Computer Music Conference*, 2000.
- [2] F. Bardet and al. Interactions geste-musique par vision artificielle. In *Reconnaissance des Formes et Intelligence Artificielle*, 2004.
- [3] C. Cadoz. Informatique et outil de création musicale. *Marsyas*, (7) :18–29, 1988.
- [4] C. Cadoz. Continuum énergétique du geste au son, simulation multisensorielle interactive d’objets physiques. *Interfaces homme-machine et création musicale*, Hermès Science Publications, pages 165–181, 1999.
- [5] C. Cadoz. Les nouveaux gestes de la musique. *Parenthèses*, 1999.
- [6] T. Coduys and C. Henry. Nouveaux matériels de captation haute precision. In *Journées d’informatique musicale*, 2003.
- [7] T. Cont, A. Coduys and C. Henry. Real-time gesture mapping in pd environment using neural networks. In *New Interfaces for Musical Expression*, 2004.
- [8] E. Fléty and M. Sirguy. Eobody : a follow-up to atomic pro’s technology. In *New Interfaces for Musical Expression*, 2003.
- [9] C. Havel and M. Desainte-Catherine. Modeling an air percussion for composition and performance. In *New Interfaces for Musical Expression*, 2004.
- [10] M. Mathews. Current directions in computer music research. *MIT Press Cambridge*, 1991.
- [11] M. Puckette. The patcher. In *International Computer Music Conference*, 1988.
- [12] M. Puckette. A real-time operating system for computer music. In *International Computer Music Conference*, 1990.
- [13] M. Puckette. Pure data : another integrated computer music environment. In *International Computer Music Conference*, 1996.
- [14] M. Wanderley. *Interaction Musicien-Instrument : Application au Contrôle Gestuel de la synthèse sonore*. Thèse de doctorat, Université Paris 6, juin 2001.
- [15] M. Wanderley and M. Battier. *Trends in Gestural Control of Music*. IRCAM, 2000.
- [16] D. Zicarelli. An extensible real-time signal processing environment for max. In *International Computer Music Conference*, 1998.