

Modélisation de sons bruités par la Synthèse Granulaire

Charles BASCOU

Mémoire de stage de DEA ATIAM
année 2003-2004

Faculté des sciences de Luminy
Université AIX-MARSEILLE II

Effectué au Centre National de Création Musicale **GMEM**
(Groupe de Musique Expérimentale de Marseille)
sous la direction de Laurent POTTIER

Résumé

Ce rapport expose le travail que j'ai effectué au GMEM (Groupe de Musique Expérimentale de Marseille) dans le cadre d'un stage de DEA ATIAM (Acoustique, Traitement du signal, Informatique Appliqués à la Musique), formation que j'ai suivie à l'IRCAM (Institut de Recherche et de Coordination Acoustique et Musique).

Ce travail a porté principalement sur la synthèse granulaire et les moyens et méthodes de produire, par celle-ci, des sons bruités environnementaux. Le but est ici de se baser sur l'analyse préalable de sons réels notamment par l'extraction de leurs propriétés stochastiques, propriétés qui se devront d'être applicables aux paramètres de la synthèse granulaire. En cela nous proposons les premières expérimentations pour l'élaboration d'un outil d'analyse/synthèse adapté aux sons bruités.

Remerciements

Je tiens à remercier toute l'équipe du GMEM au sein de laquelle cela a été un réel plaisir de travailler et plus particulièrement :

- Raphaël De Vivo pour m'avoir accueilli au sein du GMEM
- Laurent Pottier pour son aide et ses conseils tout au long du stage
- Jérôme Decque pour son assistance technique
- Jean Francois Oliver, Leopold Frey et Sophie Colas pour avoir contribué à la bonne ambiance générale

Table des matières

1	Introduction	6
2	La synthèse granulaire	8
2.1	Historique	8
2.2	Principe et type de synthèse	11
2.3	Le projet GMU (GMEM Microsound Universe)	15
2.4	Application aux sons bruités	16
3	Méthodes d'analyse	19
3.1	Analyse par segmentation temporelle	19
3.1.1	Motivations	19
3.1.2	Segmentation	20
3.1.3	Extraction de paramètres	26
3.1.4	Critique	30
3.2	Analyse Matching Pursuit	32
3.2.1	Motivations	32
3.2.2	Décomposition Matching Pursuit	33
3.2.3	Etude de cas	37
3.3	Analyse Spectral Matching Pursuit	44
3.3.1	Motivations	44
3.3.2	L'algorithme Spectral Matching Pursuit	46
3.3.3	Etude de cas	49
4	Synthèse dans MAX/MSP	56
4.1	Générateur de nombre aléatoire de loi arbitraire	56
4.2	Objet de synthèse granulaire avec enveloppe sous forme de table	58
4.3	Vers une interface graphique de contrôle...	61
5	Conclusion et perspectives	62
5.1	Bilan	62
5.2	Optimisations et Perspectives	62
5.3	Apports Personnels	63

Table des figures

1	Exemple de grain avec enveloppe à segments modulée par une forme d'onde sinusoïdale.	11
2	Schéma de l'algorithme de segmentation	20
3	Réponse en fréquence du filtre "demi-Hanning"	21
4	Réponse en fréquence du filtre sinc-hanning	22
5	Exemple d'enveloppe extraite respectivement par filtre demihanning de taille $N = 8200$ et par filtre sinc-hanning de taille $N = 8200$ et de fréquence de coupure $f_l = 10Hz$ ($f_c \simeq 20Hz$).	23
6	Exemple de détection d'attaque respectivement par seuillage constant et par seuillage dynamique.	25
7	Spectrogramme pris entre 0 et 10000 Hz de l'extrait de Turenas de John Chowning.	27
8	Resultat de l'estimation du pitch par analyse <i>pic</i>	28
9	Histogrammes des fréquences correspondant à la troisième partie de l'extrait.	29
10	Spectrogramme du son d'eau analysé d'une durée de 1.8 secondes pris entre 0 et 10000 Hz	37
11	Erreurs d'approximation en fonction du nombre d'iterations respectivement pour les atomes de Gabor et pour les atomes sinusoides amorties. Le troisième graphe représente la différence de l'erreur Gabor à l'erreur sinusoides amortie	38
12	Résultat de la décomposition Matching Pursuit du bruit d'eau entre 0 et 10000 Hz	39
13	Histogrammes normalisés des paramètres des grains détectés, respectivement d'amplitude, de fréquence en Hz et de durée en seconde.	40
14	Spectre du son d'eau resynthétisé avec enveloppe à segment exponentiel	42
15	Spectrogramme du son de grattement analysé d'une durée de 6 secondes pris entre 0 et 3000 Hz.	49
16	Spectrogramme du grain choisi pour la decomposition, d'une durée de 40 ms, pris entre 0 et 3000 Hz.	50
17	Fonction de corrélation entre le grain et le signal de départ.	50
18	Spectres residuels et fonctions de corrélation à la $10^{ième}$, $50^{ième}$ et $100^{ième}$ itération de l'algorithme.	51
19	Temps et amplitude des grains détectés.	52
20	Spectrogramme entre 0 et 3000 Hz du son resynthétisé.	53
21	Facteur de transposition correspondant aux grains détectés.	54

22	Spectrogramme entre 0 et 3000 Hz du son resynthétisé suite à une décomposition sur plusieurs grains.	55
23	Patch MAX pour la génération de nombres aléatoires de loi arbitraire sous forme messages.	57
24	Objet externe pour la génération de nombres aléatoires de loi arbitraire sous forme signal.	58
25	Objets MAX/MSP de synthèse granulaire	59
26	Patch d'aide de l'objet <i>bufGran~</i>	60
27	Interface de contrôle de l'objet <i>bufGranul~</i> de synthèse granulaire.	61

1 Introduction

Les développements connus de l'informatique musicale au cours des vingt dernières années, notamment facilités par l'envolée des puissances de calcul disponibles, mettent aujourd'hui la machine au premier plan de la création artistique.

Le GMEM est un centre de création, de diffusion et de recherche musicale, créé en 1972 par l'initiative d'un groupe de compositeurs (notamment Georges Boeuf) issu de la classe d'électroacoustique du CNRM¹. En 1996, sur décision du Ministre de la Culture, il prend le titre de "Centre National de Création Musicale".

De par son appellation même de "Groupe de Musique Expérimentale de Marseille", le GMEM a une vocation expérimentale. La recherche musicale y a toujours été liée à des problématiques de création. Le GMEM accueille des compositeurs en résidence avec la volonté de leur fournir des outils adaptés à leur demande, les accompagnant dans la concrétisation de leur projet. Ces considérations sont la base du processus qui induit la recherche et le développement au GMEM, avec principalement trois éléments constitutifs que sont :

1. l'espace et la spatialisation
2. le geste instrumental et les interfaces gestuelles
3. le matériau sonore, sa production et ses traitements

Nous allons approfondir dans ce rapport particulièrement le troisième domaine, incarné par le jeune projet de recherche GMU , ou "GMEM Microsound Universe", consacré à la synthèse microsonore. En effet le GMEM est attaché à pouvoir proposer aux musiciens des outils innovants pour la production sonore. La synthèse microsonore (ou granulaire), notamment par son contrôle non trivial, constitue un axe de recherche fécond dans cette perspective. Ce type de synthèse n'est pas nouveau, ayant été largement étudié par des chercheurs-musiciens comme Iannis Xenakis, Curtis Roads ou Barry Truax. Mais il offre toutefois un vaste champ encore peu exploré dans son application à la resynthèse de sons naturels. Ses principes intrinsèques le prédisposent tout particulièrement aux sons bruités. La sensibilité du GMEM est d'ailleurs attaché au mouvement "Musique Concrète" initié par Pierre Schaeffer au milieu du *XXième* siècle, mouvement qui amenait au statut musical les sons quotidiens environnementaux. De là, est née la volonté de proposer des outils permettant la synthèse et la manipulation de ce type de sons.

¹Conservatoire National de Région de Marseille

Nous allons, dans ce rapport, tout d'abord rappeler l'histoire et les principes de la synthèse granulaire, ceci nous permettant d'affiner la définition du travail qu'il m'était demandé de réaliser. Dans une deuxième partie, nous détaillerons les méthodes d'analyse envisagées, sous-tendues par un tel projet. Enfin, la troisième partie exposera brièvement le développement impliqué pour la synthèse dans l'environnement MAX/MSP.

2 La synthèse granulaire

2.1 Historique

Norbert Wiener La notion de quantum sonore apparaît avec N. Wiener en 1925 qui lors d'une conférence sur la physique quantique utilise comme référence la musique. Il donne des détails sur la relation entre temps et fréquence, il énonce que la précision dans le domaine temporel entraîne des imprécisions dans le domaine fréquentiel et vice-versa. Apparaît alors une notion de quantum sonore dès lors qu'on connaît la taille minimum d'une particule sonore (en fonction de sa fréquence) d'où l'idée, similaire à celle de la mécanique quantique qui décrit la matière, de pouvoir décrire les sons en termes de taille et nombre de grains.

Denis Gabor En 1946 le physicien prix Nobel D. Gabor fabrique une machine utilisant un système de grains pour reproduire des sons. Il argumente ses découvertes par un article appelé «Theory of communication» suivi un an après par «Acoustical Quanta and the Theory of Hearing». Il y parle notamment des problèmes de l'analyse de Fourier. Gabor dit que alors que les mathématiques y sont absolument correctes, cette théorie ne convient pas à la description des sons notamment ceux à fréquences variables. Un autre problème est qu'utiliser des ondes sinusoïdales implique que la durée du signal est infinie. Gabor présente l'idée d'utiliser les théories de Physique quantique pour l'étude des signaux sonores.

Gabor effectue des études sur les seuils de discrimination. Pour cela il utilise les résultats des expériences menées par Buerck, Kotowski, Lichte puis plus tard par Shower et Biddulph. Il examine la durée nécessaire d'un son pour reconnaître sa hauteur à différentes fréquences. Ses études montrent un certain nombre de choses :

- Les fréquences entre 500Hz et 1000Hz doivent être jouées au minimum pendant 10 ms avant d'être perçues comme des hauteurs.
- Si une fréquence change, il faut 2 fois le temps qu'il a fallu pour entendre la première fréquence pour discerner le changement. Cette durée de reconnaissance de changement varie avec la fréquence. A basses fréquences, l'homme discerne plus vite.
- Le discernement d'un changement d'amplitude prend minimum 21 ms.

Gabor émet alors une théorie mathématique sur les surfaces de seuil de perception. Il utilise ces idées pour développer un certain nombre de machines. Le *Tempophon* fabriqué par la compagnie allemande *Springer* est une machine à changer le pitch ou la durée de documents sonores et est directement issue d'une machine de Gabor.

Abraham Moles Les recherches de Moles sur la théorie de l'information l'ont amené à la conclusion suivante : pour pouvoir définir et déchiffrer un message, qu'il soit visuel ou sonore, on doit prendre en considération les caractéristiques psychophysiologiques du récepteur, car elles seront circonstancielles. «Un message est un groupe fini, ordonné, d'éléments de perception puisés dans un répertoire et assemblés en une structure. Les éléments de ce répertoire sont définis par les propriétés du récepteur. »

Moles est le premier à dire que le nombre de sensations que reçoivent nos organes psychophysiologiques est quantifiable. Moles a étudié le pouvoir de l'ouïe pour résoudre des petites différences de fréquence et d'amplitude et parle d'un répertoire d'éléments audibles qui est plus ou moins limité à 340000. Moles représente «l'atome sonore» comme une cellule tridimensionnelle ayant pour côté le seuil différentiel de fréquence $\Delta F/F$, le seuil différentiel d'intensité $\Delta L/L$, le seuil différentiel de durée $\Delta T/T$. Cette formulation du quantum sonore est à peu près la même que celle de Gabor mais est plus précise car grâce à elle on peut répertorier le nombre existant de «quanta de sensation».

Iannis Xenakis Xenakis a été le premier musicien à créer une théorie structurale de la composition où les grains sonores sont les éléments constitutifs de base. Avant d'utiliser ces idées dans la musique électronique, il réalise des pièces instrumentales basées sur le concept de cellule sonore dont les paramètres sont régis par des lois statistiques. Dans *Pithoprakta* pour orchestre à cordes écrit en 1955-56, il utilise le glissando comme cellule élémentaire sonore, les distribuant librement en fréquence. Dans les années 60, il était le premier à s'intéresser à la "Theory of communication" de Gabor et à l'utiliser en composition. Xenakis soulignait la propriété triple de chaque grain : durée, fréquence et intensité. Il avançait que toute musique était composée d'un grand nombre de grains élémentaires. D'un point de vue compositionnel, il ajoutait une troisième dimension à l'espace temps-fréquence usuel, définissant le grain. Il invite ainsi le compositeur à penser son oeuvre avec d'autre forme d'abstraction du son que celle conventionnellement utilisée.

Curtis Roads En 1975, C. Roads essaie de vérifier et d'expérimenter les théories du quantum sonore de Gabor. Il construit un système informatique automatisé pour la génération de grains sonores synthétiques à l'Université de Californie à San Diego, en utilisant le système informatique *Borroughs B6700*, avec le programme *Music V*. Dans ce système, chaque grain a une durée et une fonction d'amplitude fixes pendant que les fréquences, formes d'onde et amplitudes sont variables. Un paramètre très important est la forme de

l'enveloppe du grain, il utilise une enveloppe qui combine la courbe de Gauss suggérée par Gabor et le rectangle suggéré par Xenakis obtenant un début et une chute légère et en même temps un sommet soutenu. Ce système pouvait générer 32 grains simultanés et des densités de 1600 grains par seconde. Roads est l'auteur de nombreux ouvrages dédiés à la synthèse granulaire dont [15],[14].

Barry Truax B. Truax dans les années 80 construit son propre système de synthèse granulaire, installé à l'Université de SFU au Canada et au GMEB de Bourges. La diffusion de ses travaux musicaux et de recherche a contribué à l'intérêt de la communauté scientifique et musicale pour cette forme de synthèse dans les années 90 qui ont vu l'enrichissement des techniques granulaires. Il a notamment mis à jour les possibilités de changement d'échelle temporelle et de transposition en temps réel de sons enregistrés dans [17].

2.2 Principe et type de synthèse

Suite à ce bref historique des recherches et trouvailles de pionniers de la synthèse granulaire, nous allons en présenter maintenant le principe général notamment par l'exposé des diverses lois et paramètres les plus usitées. Nous poursuivrons par les différentes formes que peut prendre ce type de synthèse et leurs applications spécifiques.

A la base d'une texture granulaire se trouve le grain sonore, extension de la notion de quanta sonore introduite par Gabor. La synthèse granulaire permet, par l'amas de ces grains selon des stratégies diverses, la génération de textures sonores (au sens le plus général) complexes. On peut distinguer plusieurs niveaux de paramètres régissant cette synthèse. Fondamentalement, on manipule les paramètres du grain sonore lui-même, comme sa longueur ou son amplitude. A un niveau supérieur, on élabore des stratégies de contrôle du flux de grains, ou comment organiser le déclenchement, l'évolution des paramètres de chacun des grains constituant le flux.

Paramètres du grain Un grain est constitué d'une forme d'onde et d'une enveloppe modulante. La réunion des paramètres de ces deux entités va définir de manière unique notre grain.

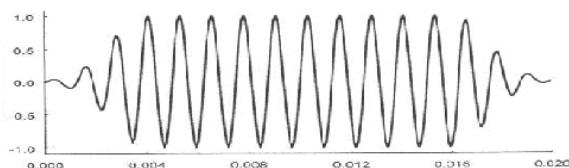


FIG. 1 – Exemple de grain avec enveloppe à segments modulée par une forme d'onde sinusoïdale.

La forme d'onde prise comme base pour la construction d'un grain peut être de nature très diverse et le principe même de la synthèse granulaire n'en restreint pas le choix. On peut distinguer tout de même des grands principes dans le choix de la forme d'onde qui sont :

- *onde sinusoïdale* : forme d'onde la plus simple du point de vue spectral permettant, moyennant des contrôle de haut niveau, la mise en oeuvre de structures complexes.

- *échantillon sonore* : on lit ici la forme d'onde dans une table contenant un son numérisé. Ceci est à la base de techniques de type *time stretch* ou *pitch shift* consistant à décorrélérer les paramètres de hauteurs et de durée des sons enregistrés.
- *onde synthétique complexe* : d'autres formes d'onde synthétiques complexes ont été introduites par les pionniers de la synthèse granulaire. Ceci incluant les forme d'ondes générées par synthèse FM, par distorsion non linéaire, par somme de sinus (Forme d'onde Formantique),...

Bien que chacun des cas ci-dessus possède leurs propres paramètres, le *pitch* de la forme d'onde est une donnée qui reste commune à tous et en cela constitue un paramètre essentiel de la synthèse granulaire. Il a un rôle très important dans la construction de structures sonores complexes. C'est en effet un des paramètres les plus probants perceptuellement et qui a été l'un des premiers dont le contrôle a été appréhendé d'un point de vue compositionnel(Xenakis). Notons que le terme *pitch* n'est pas réellement adéquat dans le cas de l'échantillon sonore, devenant un paramètre relatif à la hauteur de l'échantillon lui même. On pourra alors lui préférer le terme de transposition. Un autre paramètre commun à ces trois catégories est le paramètre de phase. Même si historiquement, il est peu utilisé dans le cas d'onde synthétique, il n'en reste pas moins essentiel pour définir le point de départ de lecture d'un échantillon sonore. On prendra alors pour le caractériser le terme plus évocateur de *begin*.

L'enveloppe modulante est la deuxième entité centrale d'un grain sonore. Elle détermine entre autres les paramètres importants que sont l'amplitude et la taille du grain. Comme pour la forme d'onde, les paramètres contrôlables sont directement liés au type d'enveloppe choisi qui sont principalement :

- *enveloppe à segment* : généralement composé de trois phases que sont l'attaque, la partie tenue et la chute. Ces trois phases sont principalement paramétrées par leurs durées respectives, respectivement le paramètre d'*attack*, de *steady state* et de *release*. Leur somme va ainsi déterminer la durée totale du grain. On peut noter que même si les segments sont traditionnellement linéaires, des variantes peuvent exister comme l'utilisation de segment exponentiel ou demi-cosinus pour les phases d'attaque et de chute.
- *enveloppe fonctionnelle* : générées lors de la synthèse, elles suivent

une fonction que l'on aura préalablement déterminée. C'est le cas par exemple des enveloppes "hanning" ou gaussiennes. Chaque fonction aura ses propres paramètres qui pourront influencer entre autre sur la durée ou la puissance de l'enveloppe.

- *enveloppe basée sur une table* : l'enveloppe est définie sous la forme d'une table d'onde. On peut ainsi utiliser des enveloppes arbitraires dessinées ou calculées, mais dont la forme ne pourra être paramétrée pour chaque grain. On peut toutefois modifier la durée de l'enveloppe en lisant plus ou moins rapidement la table d'onde. Ce type d'enveloppe a donc un paramètre de taille *length*.

Le paramètre de longueur de grain est très important d'un point de vue perceptif (hauteur et amplitude perçue du grain,...) . En effet, il va modifier l'étalement spectral de l'enveloppe et par convolution celui de la forme d'onde. Ainsi, autant nous sommes capable, avec une longueur de 100 ms, de bien discerner le contenu spectral de la forme d'onde, autant des longueurs de moins de 5 ms ne produisent que des "clicks" colorés. Un des paramètres communs essentiels de l'enveloppe modulante est bien évidemment son amplitude *amp* qui influera sur l'amplitude même du grain.

On peut affecter à l'enveloppe le paramètre de spatialisation *pan* qui permet de distribuer les grains sur plusieurs localisations spatiales. Ceci peut en effet se faire en spécifiant une amplitude d'enveloppe différente pour chacun des canaux de sortie audio. Par ailleurs, on pourra noter que l'utilisation d'enveloppes fonctionnelles précalculées dans des tables est souvent préférable à un calcul en temps réel de celles ci, surtout dans le cas de fonctions dont le seul paramètre est l'étalement temporel (enveloppe hanning).

Paramètres du flux Nous avons ainsi défini les paramètres les plus importants des entités sonores en jeu dans un tel type de synthèse. Mais, l'intérêt serait bien maigre sans l'ajout de stratégies de contrôle de tous ces paramètres. Ne serait-ce que pour la question du déclenchement de ces grains, il n'est pas question ici de les déclencher un à un à la manière d'un échantillonneur mais plutôt de gérer leur distribution temporelle par des stratégies de plus haut niveau. Nous avons donc généralement un paramètre d'espacement temporel entre le déclenchement de grains successifs, qui peut selon les applications être sous la forme d'une fréquence ou d'un temps d'espacement (période). Ce paramètre va directement influencer, avec la durée du grain, sur la densité de grains, autrement dit le nombre de grains jouant simultanément. La fréquence de déclenchement a des effets bien particuliers

selon l'ordre de grandeur dans laquelle elle évolue. Cela a à voir avec le principe de fusion psychoacoustique des périodicités. En effet, à fréquence basse on obtiendra un effet rythmique qui se transformera en hauteur perceptible avec son accroissement. Ceci introduit par ailleurs plusieurs types de synthèses selon les propriétés de la fréquence de déclenchement :

- *Pitch Synchronous Granular Synthesis* : les grains sont tous identiquement espacés ou selon une relation linéaire. Cette technique est utilisée notamment pour les méthodes de synthèse vocale par forme d'onde formantique comme dans le programme CHANT de Xavier Rodet [2]. L'intérêt de la PSGS est, en périodisant le grain, d'avoir la propriété d'échantillonner le spectre de ce dernier dans le domaine fréquentiel. On synthétise donc un signal harmonique, ayant pour fondamentale la fréquence de déclenchement et pour enveloppe spectrale le spectre du grain. Pour la synthèse vocale, les grains FoF sont des sommes de sinusoides fenêtrées correspondant aux formants des parties voisées.
- *Asynchronous Granular Synthesis* : Les grains sont déclenchés de manière aléatoire. Ceci est utilisé pour la création de textures sonores de type “Nuages” dans l'espace temps-fréquence initialement introduites par Xenakis. L'espacement des grains peut être compris entre deux bornes ou choisi parmi une liste finie de valeurs.

La génération d'un grand nombre de grains implique le difficile contrôle de chaque grain particulier. Il est de plus intéressant d'avoir une bonne variabilité des paramètres pour créer des structures complexes. On introduit ainsi un degré d'aléatoire ou *randomness* dans la variation des paramètres. Nous avons donc la possibilité d'ajouter à chaque paramètre un générateur de nombre aléatoire régi par un fonctionnement spécifique. La méthode la plus usitée est la génération de variables équiprobables entre deux bornes. Mais, on peut aussi contrôler selon des lois plus complexes comme choisir dans une liste finie de valeurs avec pour chacune une probabilité particulière.

Curtis Roads dans [15] a rappelé toutes les diverses variantes d'appellation selon les contraintes imposées au schéma général de synthèse. Leur liste est exhaustive... On pourra citer la “Wavelet Synthesis” qui corrèle le pitch est la durée du grain, les différentes variantes de la “Glisson synthesis” qui travaille avec des grains de type glissandi. D'autres types de synthèse s'attache à définir et manipuler des macrostructures granulaires comme la “Trainlet synthesis” qui génèrent des groupes de grains de même hauteur. Toutes ces dénominations ne sont là que pour spécialiser le caractère très général de la synthèse granulaire.

2.3 Le projet GMU (GMEM Microsound Universe)

Ce projet a démarré l'année dernière, c'est un nouveau programme de recherche à long terme lancé par le GMEM. Il vise à mettre au point un système intégré permettant la production et le contrôle de sons de synthèse pour des applications musicales.

Plusieurs points caractérisent ce système :

- Les techniques de synthèse choisies se démarquent des techniques de synthèse classiques généralement basées sur des oscillateurs continus. On privilégie ici le grain sonore et non le partiel sonore. Ces techniques sont destinées à la production de sons qui ne sont pas nécessairement harmoniques et doivent permettre d'explorer un domaine beaucoup plus large, plus proche des domaines de la musique concrète à laquelle la sensibilité du GMEM est rattachée.
- Le système est prévu pour intégrer aussi des techniques de contrôle appropriées, des outils de transformation du timbre et de spatialisation. Les variations des paramètres de synthèse peuvent induire des modifications corrélées des paramètres de traitement et de spatialisation.
- Les paramètres de contrôle de bas niveau seront regroupés et contrôlés par des paramètres perceptifs et continus (interpolations), permettant aisément aux compositeurs de créer de nouveaux types de sons.
- Le système sera opérationnel en temps-réel (environnement MAX/MSP)
- Des techniques d'analyse de sons, d'extraction de paramètres, fonctions de distribution statistique permettront de configurer automatiquement les différents modules du système.
- Le système pourra mener à terme à une typologie et une paramétrisation de sons bruités et environnementaux permettant au musicien d'explorer de nouveaux espaces sonores.

Mon travail s'est inscrit dans ce projet avec une focalisation particulière sur l'analyse et la synthèse de sons bruités par techniques granulaires. En effet, des outils de synthèse sont opérationnels grâce notamment au travail de Loic Kessous. Jean Francois Oliver a de plus élaboré l'année dernière des méthodes de contrôle avancées dans MAX/MSP et de traitement de paramètres dans OpenMusic. La partie d'analyse-resynthèse granulaire com-

mençait donc avec le début du stage.

2.4 Application aux sons bruités

Nous allons nous intéresser principalement ici aux sons bruités et aux possibilités qu'a la synthèse granulaire de les représenter. Tout d'abord qu'appelle-t-on son bruité? Il ne s'agit pas, comme on pourrait être poussé à le croire, de n'importe quel son contenant une composante aléatoire non-négligeable, la synthèse granulaire n'ayant dans ce cas qu'un bien maigre intérêt comparé aux méthodes de synthèse additive sinusoides+bruits par exemple. Tout d'abord, les sons bruités n'ont pas nécessairement une hauteur perceptible. Mais, une de leur propriété importante est d'avoir une structure temporelle spécifique, régissant leur évolution (et leur répétition) dans le temps. A la différence de méthodes de type additive, on souhaite paramétriser les sons autant d'un point de vue microscopique, avec par exemple leur composition spectrale locale, que macroscopique, représentant leur structure temporelle de manière compacte, ceci étant dans le but d'une application en synthèse granulaire. Nous nous intéressons donc ici principalement au son ayant une "cohérence granulaire" permettant une telle application. Nous nous devons d'explicitier ces propos. On voudrait représenter des sons dont la structure peut être vue comme la superposition de grains plus ou moins complexes. On étend ainsi la notion de quanta sonore de Gabor, qui avance la possible représentation de tout type de sons par une accumulation de grains élémentaires. Ici le grain garde un sens particulier et est directement lié au son que l'on souhaite représenter. Prenons l'exemple d'un bruit de pas sur un tas de feuille morte. On peut le voir comme l'accumulation de "microsons" de craquements de feuille avec une structure temporelle et une variabilité spectrale spécifique. Verser du sable sur une plaque de fer produit de la même manière la superposition denses de "tick" élémentaires. Il existe en effet de nombreux exemples de son rentrant dans cette classe. En fait, dès lors que l'on a, dans le monde physique, plusieurs occurrences d'une même structure, d'un même phénomène on peut s'attendre à avoir ce type de sons. On citera les sons de grattements, de frottements, de craquements, de bruissements,... On remarquera que cette catégorisation est directement liée à l'action effective, n'induisant à priori rien sur leurs propriétés sonores. Pourtant, nous avons incorporé la variabilité de ces sons causaux étant à même de leur associer avec plus ou moins d'assurance une catégorie. Il y a donc bien des paramètres influant de ces sons justifiant une catégorisation paramétrique par analyse.

Le but est donc ici d'analyser des sons bruités réels pour en tirer des paramètres applicable en synthèse granulaire qui puisse résulter en un son at-

tribuable à la classe du son de départ. Il nous faut donc pouvoir détecter les paramètres des grains propres au son analysé et d'en paramétrer l'évolution dans le temps. Nous sommes parti du postulat que ces évolutions pouvaient être représentées par des distributions stochastiques. A terme cela doit aboutir sur une possible classification des sons bruités selon leur caractéristiques, permettant ainsi en synthèse divers traitements comme la création de sons hybrides ou plus simplement un contrôle du son bruité par des paramètres en corrélation avec le monde physique.

Existant Nous n'avons pu que constater la relative rareté des travaux existants sur la synthèse granulaire en relation avec les sons bruités. En effet, il y a bien des approches empiriques pour tenter de synthétiser de telles classes de sons mais pas réellement de méthodes basées sur une analyse de cas réels. Pourtant plusieurs travaux nous ont motivé dans l'approfondissement d'une telle approche.

Tout d'abord, les résultats des travaux de Damian Keller musicien-chercheur au CCRMA de Stanford qui a notamment travaillé avec Barry Truax, sont intéressants sous divers aspects. Il a en effet travaillé sur les "modèles écologiques" de sons appliqués à la synthèse granulaire [9] [18], qui tente de reproduire des sons bruités environnementaux. Il a toutefois une approche plus phénoménologique pour la modélisation se basant sur les propriétés physiques des actions effectrices. Il travaille en parallèle sur la perception notamment par l'étude des invariants qui nous permettent de caractériser de tels types de sons. Il a par exemple modélisé, par l'étude de son mouvement, l'évolution des paramètres granulaires du son d'une balle rebondissante. Il a aussi des résultats intéressants sur des sons de brisures de vitres, distribuant plusieurs grains caractéristiques naturels aléatoirement. Les sons obtenus ont une connotation très naturelle, cela nous motivant sur la capacité qu'a la synthèse granulaire dans son application aux sons bruités. Mais nous ne pouvons utiliser de tels travaux, l'approche attendue ici étant de l'ordre de l'analyse-synthèse prenant pour base uniquement le son.

D'autre part, certains pans de la recherche en traitement du signal sonore, même si leur application n'est pas directement lié à la synthèse granulaire, peuvent être intéressants pour nous. Ainsi, les travaux de Pierre Hanna et Myriam Desainte-Catherine sur la synthèse de bruits par somme de sinusoides donne des résultats encourageant pour notre modélisation [7] [6] . Ils s'attachent à synthétiser des bruits colorés par transformée de Fourier inverse, dont l'application directe est dans l'analyse-transformation-synthèse par TF discrète. Ils ont montré qu'il était possible de générer du bruit coloré en sommant un nombre fini de sinusoides dont les fréquences et phases sont choisies

aléatoirement pour chaque fenêtre de resynthèse. On peut alors voir une fenêtre comme une superposition de grains sinusoidaux, nous montrant ainsi la possibilité de générer en synthèse granulaire des bruits colorés, la couleur du bruit étant en relation avec la distribution statistique des fréquences centrales des grains. Selon ce principe, on pourrait imaginer extraire la densité spectrale de puissance d'un bruit permettant ensuite la resynthèse granulaire. Cela peut être intéressant mais ce n'est pas le propos ici. Cela a néanmoins le mérite de nous conforter dans l'idée de pouvoir représenter du bruit par une accumulation de grains sonores simples.

Suite à ce constat sur la relative pauvreté des travaux antérieurs sur cette problématique, il a fallu tout d'abord chercher des méthodes d'analyse permettant la décomposition du son de départ en grains, desquels on pouvaient tirer des paramètres exploitables. Le prochain chapitre expose les diverses méthodes que nous avons évaluées pour réaliser cette tâche.

3 Méthodes d'analyse

Cette partie expose les différentes pistes que nous avons explorées pour l'analyse granulaire de sons bruités. Les deux premières, que sont la segmentation par détection d'attaque et l'algorithme Matching Pursuit, sont des techniques existantes, plus ou moins utilisées en traitement du signal, et dont nous proposons une application à l'analyse granulaire. Nous introduirons enfin une troisième méthode, que nous nommerons Spectral Matching Pursuit, extension de la méthode Matching Pursuit traditionnelle au domaine fréquentiel qui se trouve particulièrement prometteuse pour la décomposition de signaux en grain sonore arbitraire.

3.1 Analyse par segmentation temporelle

3.1.1 Motivations

Comme nous l'avons vu plus haut, le but de la phase d'analyse est de tenter de décomposer le son source en microentités sonores, desquelles on pourrait tirer des paramètres pertinents pour la resynthèse. Une des premières idées qui vient naturellement à l'esprit est d'utiliser des techniques de segmentation dans le domaine temporel pour isoler les grains et ainsi pouvoir en tirer des paramètres propres à chacun. Nous nous sommes pour cela inspirés des méthodes de détection d'attaque dont le champ d'application va de la détection de tempo, au suivi de partition ou à la reconnaissance d'instrument. On part ici du postulat que chaque grain peut être d'une part détecté par les propriétés dynamiques du signal et d'autre part qu'il peut être isolé du point de vue temporel pour une extraction précise de ces paramètres. Ainsi si x est le signal analysé et $D = \{g_i\}$ un ensemble de grains, cette méthode s'applique à tout signal x de longueur N tel que :

$$x[k] = \sum_{i \in D} \alpha_i g_i[k] \quad \text{avec} \quad \forall i, j \in D^2 \quad \sum_{k=0}^N g_i[k]^2 \cdot g_j[k]^2 = 0$$

Autrement dit, la segmentation ainsi effectuée par cette méthode ne peut décomposer que des sons dont la microstructure granulaire ne se recouvre pas dans le temps. En réalité il peut être possible de détecter des attaques de grains se recouvrant mais l'analyse des paramètres de ceux ci devient beaucoup plus problématique. Nous nous restreindrons pour l'instant à cette seule application.

Le mode opératoire pour cette méthode d'analyse est tout d'abord la segmentation du signal par détection de variations brutales dans sa dynamique,

puis l'extraction des paramètres propres à chaque segment suivi éventuellement d'un traitement de ces paramètres pour les rendre exploitables par la synthèse.

3.1.2 Segmentation

Les algorithmes de détection d'attaque sont devenus très populaires dans de divers domaines de traitement de signal. Pour notre application nous nous sommes basés sur une méthode de segmentation issue d'un algorithme de détection de tempo de morceaux musicaux proposée par Scheirer dans [16]. Cette méthode extrait en premier lieu l'enveloppe d'amplitude du signal et en détecte les variations brusques. L'algorithme peut se schématiser comme suit :



FIG. 2 – Schéma de l'algorithme de segmentation

Cette méthode d'extraction d'enveloppe porte le nom de “rectify and smooth method”. Il s'agit dans un premier temps de rectifier le signal analysé x c'est à dire calculer $rx = x^2$, ceci permettant d'avoir un signal qui oscille autour de sa fonction d'enveloppe. Le filtrage passe bas suivant ne gardera que la composante évoluant lentement au cours du temps qui représente une bonne estimation de l'enveloppe d'amplitude du signal source. La fonction d'enveloppe est alors différenciée pour en détecter les variations positives brusques par seuillage.

Choix du filtre Le choix du filtre est très important pour bien séparer ce qui dans le signal rectifié est de l'ordre des fréquences relatives au spectre de la forme d'onde et celles relatives à l'enveloppe d'amplitude. Scheirer propose pour le filtrage du signal rectifié de convoluer par une demi fenêtré Hanning de longueur 200 ms. ce filtre RIF a une fréquence de coupure à -3dB aux environ de 10 hertz et une pente de -6dB/octave. De part sa discontinuité en 0, ce filtre n'est pas à phase linéaire. En réalité dans notre cas, ormis la considération du délai introduit dans la bande de fréquence qui nous intéresse (environ $0 - 20Hz$), la linéarité de phase du filtre n'est pas une contrainte forte. La réponse en fréquence d'un tel filtre est représentée sur la figure 3.

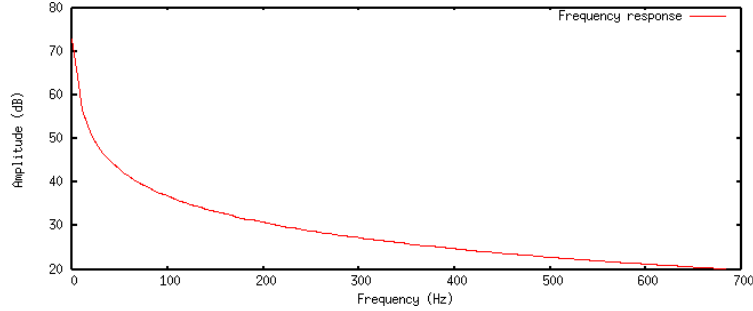


FIG. 3 – Réponse en fréquence du filtre “demi-Hanning”

Après plusieurs tests sur des divers exemples sonores notamment comportant de fortes composantes fréquentielles graves, nous avons mis à jour la limite d’utilisation d’un tel filtre pour l’extraction d’enveloppe. En effet, les composantes fréquentielles de la forme d’onde interféraient avec les variations d’amplitude du signal. Lors de la détermination des temps d’attaque par seuillage de la fonction d’enveloppe différenciée, on détecte à la fois les variations positives d’enveloppe et les périodes des composantes graves. Ceci est principalement dû à la trop faible pente de coupure du filtre (rolloff).

Cela nous a donc poussé à utiliser un autre type de filtre pour l’extraction de l’enveloppe d’amplitude du signal. Nous avons utilisé la méthode de la fenêtre pour le synthétiser. Le gain de ce filtre doit idéalement être $H(e^{2j\pi f}) = \mathbf{1}_{(-f_c, f_c)}(f)$ (avec $\mathbf{1}_{(-f_c, f_c)}(f)$ fonction de heavyside centrée en 0 et de largeur $2f_c$). La réponse impulsionnelle d’une telle fonction de transfert est :

$$h(t) = \int_{-1/2}^{1/2} H(e^{2j\pi f}) e^{2j\pi t f} df = \frac{\sin(2\pi t f_c)}{\pi t}$$

Cette réponse impulsionnelle étant infinie, nous l’avons pondéré par une fenêtre de hanning, préférable à une fenêtre rectangulaire, n’introduisant des lobes secondaires qu’à -40dB contre -13dB. Ceci est par ailleurs au détriment de l’augmentation de la largeur de bande du filtre. En effet, ayant multiplié les deux signaux, on obtient une convolution dans le domaine spectral. La bande passante du filtre devient alors :

$$B = 2f_c + B_w \quad \text{avec } B_w \text{ bande passante de la fenêtre Hanning}$$

B_w est directement liée à la taille de la fenêtre hanning utilisé et peut être calculé par :

$$B_w = \frac{C_w}{M} \cdot Fe$$

avec M taille de la fenêtre, $C_w = 1.78$ coefficient de largeur de bande passante (données pour les fenêtres usuelles) et Fe fréquence d'échantillonnage. Nous pouvons compenser, pour la synthèse du filtre, cet élargissement de bande en prenant $f_c = f_l - B_h/2$ avec f_l fréquence de coupure demandée. Il faudra toutefois $B_h < f_l$ et ainsi $M > \frac{C_w}{f_l} \cdot Fe$. En prenant par exemple $f_l = 20Hz$ et $Fe = 44100$, la taille de la fenêtre ne devra pas être inférieure à environ 4000 échantillons. De plus, diminuer la taille de la fenêtre aura pour effet de diminuer la pente de coupure du filtre. Nous avons donc dû trouver un compromis entre sélectivité en fréquence du filtre (directement liée à la longueur de la fenêtre) et coût de calcul. Pour nos tests nous avons utilisé une fenêtre de 8000 échantillons nous permettant de descendre jusqu'à une fréquence de coupure de 10 Hz. La réponse en fréquence, avec $f_c = 5Hz$ et $M = 8000$ sont représentées en figure 4.

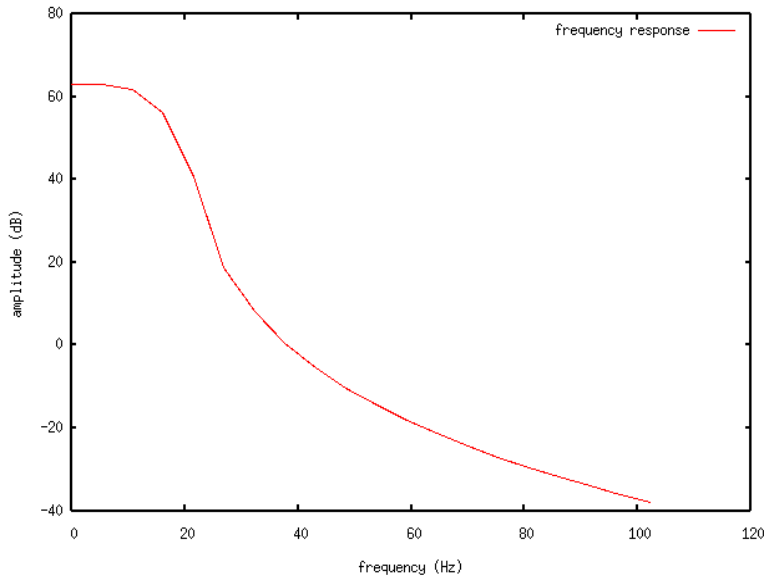


FIG. 4 – Réponse en fréquence du filtre sinc-hanning

Il est par ailleurs bon de préciser que le filtre créé n'est pas à proprement parler causal, sa réponse impulsionnelle étant centrée en 0. Pourtant, on peut le considérer comme tel en acceptant un retard de $M/2$, n'étant pas ici dans

des contraintes de temps-réel.

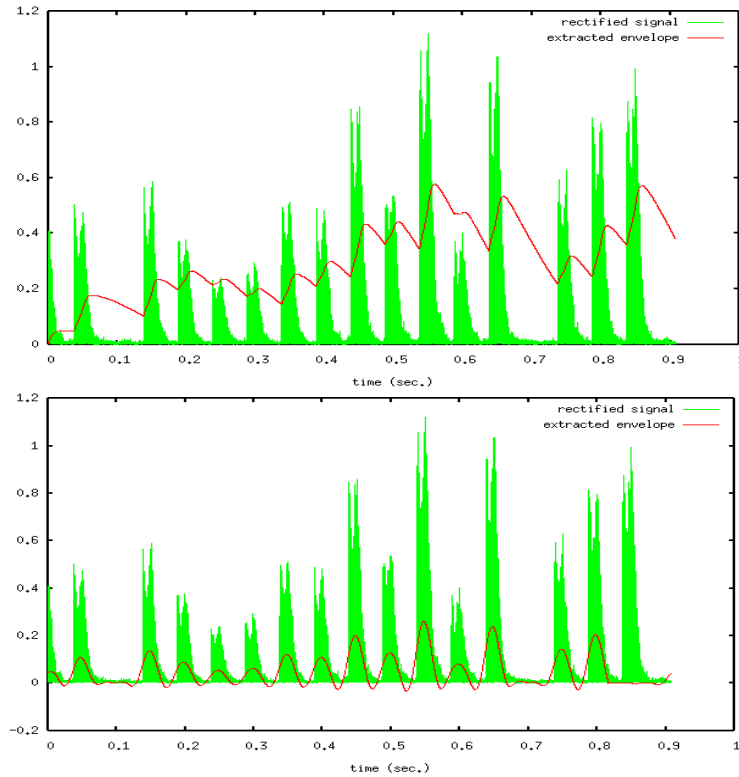


FIG. 5 – Exemple d’enveloppe extraite respectivement par filtre demi-hanning de taille $N = 8200$ et par filtre sinc-hanning de taille $N = 8200$ et de fréquence de coupure $f_l = 10Hz$ ($f_c \simeq 20Hz$).

Un exemple d’extraction d’enveloppe est représenté en figure 5. On notera tout d’abord que les signaux d’enveloppe ont été remis à l’échelle, le gain des filtres étant environ de 60dB (d’un facteur 1000) dans leur bande passante. On voit clairement que le filtre demi-hanning n’atténue pas aussi fortement les hautes fréquences que le sinc-hanning. Les transitions sont en effet beaucoup plus brutales pour le premier. De plus, le filtre demi-hanning introduit un décalage qui n’est pas significatif de l’enveloppe d’amplitude réelle du signal. Ceci est principalement dû à un trop fort gain des fréquences très faibles dans sa bande passante (fonction de transfert en pic). Ce comportement n’est pas attendu ici, le but étant de coller au plus près des variations de dynamique du signal, dès lors qu’elles sont dans la bande de fréquence spécifiée (0-10Hz). Le choix du filtre sinc-hanning semble désormais justifié,

filtre qui à coût de calcul égal, offre de bien meilleures spécifications pour l'extraction d'enveloppe appliquée à la détection d'attaque.

autres traitements Une fois l'enveloppe d'amplitude du signal extraite, on calcule sa différentielle pour détecter ses variations brusques. Ceci peut se faire avec une filtre différentiateur dont l'équation aux différences est de la forme :

$$y(t) = 0.5x(t) - 0.5x(t - 1)$$

Sur l'enveloppe ainsi différenciée, on effectue ensuite un seuillage pour détecter les attaques (ie. les plus grandes variations positives). Le choix du seuil est important selon le type de son analysé et ce que l'on veut détecter. Il doit pouvoir être ajustable par l'utilisateur. Mais il est aussi dépendant de la puissance du signal. En effet, un seuil constant pénalise trop fortement les attaques de grains faibles en intensité :

$$\text{soit } x_1(t) \text{ et } x_2(t) = 0.2 * x_1(t) \text{ on a alors } y_2(t) = 0.2 * y_1(t)$$

Par ailleurs, un seuil trop faible, qui pourrait à priori détecter des attaques d'intensités diverses, donne trop d'erreurs de dédoublement d'attaque pour les grains d'intensité forte. Ainsi, une petite variation d'enveloppe d'amplitude au sein de la chute ("decay") d'un grain peut être faussement détectée comme une attaque. Or, la méthode se doit, particulièrement dans notre cas, de détecter des attaques d'intensités diverses, le paramètre d'amplitude pour chaque grain étant une donnée potentiellement intéressante à extraire par la suite. L'idée pour améliorer la détection est alors de prendre un seuil bas, fonction des plus faibles attaques à détecter, et de pondérer ce dernier par la puissance locale du signal, évitant ainsi les artefacts de dédoublements. Notons que cela n'est valable que dans le contexte de départ, dans lequel les grains sonores constituant le son analysé ne se recouvraient pas (ou peu) dans le temps. En effet, dans le cas contraire, la donnée de puissance ne peut plus être attribuée à un seul grain, le seuil ne pouvant plus s'adapter en conséquence.

La puissance du signal est calculée sur une fenêtre de taille N suivant :

$$P_x(t) = \sum_{k=0}^N x^2(t - k)$$

La taille N pourra être choisie pour influencer la détection d'un grain par rapport au grain précédemment trouvé. Avec N grand, un grain faible peu de

temps après un grain fort ne sera pas détecté. Ce paramètre n'est toutefois pas très probant, ses effets interférant avec ceux de la bande passante du filtre d'extraction d'enveloppe. Nous avons, pour nos tests, utilisé une taille de 1024 échantillons (correspondant à une fréquence de 40 Hz), permettant à des grains ne se recouvrant pas et dont la fréquence de déclenchement n'excède pas les 20Hz de peu s'influencer les uns les autres. Un exemple de détection d'attaque par seuillage est présenté en figure 6.

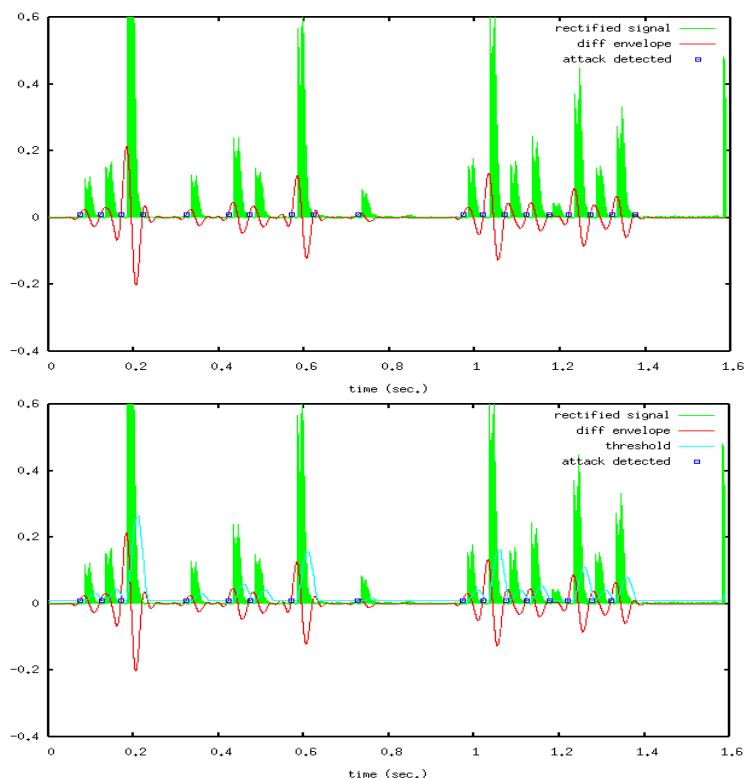


FIG. 6 – Exemple de détection d'attaque respectivement par seuillage constant et par seuillage dynamique.

Pour cet exemple, nous avons gardé la fréquence de coupure et la taille du filtre d'extraction d'enveloppe à respectivement $f_c = 20Hz$ et $N = 8200$. Le seuil est fixé à la valeur très faible de $s = 0.01s$ pour garantir la détection du grain à $t = 1.2$. On remarque que par seuillage constant l'algorithme fait trois erreurs de dédoublement à $t \simeq 0.2$, $t \simeq 0.6$ et $t \simeq 1.4$. Ces trois erreurs sont dues au fait que, bien que cela soit peu visible sur la figure, les grains constituant le son analysé ont une enveloppe d'amplitude comportant plusieurs segments. Certaines phases ascendantes de l'enveloppe sont

ainsi détectées, quand l'intensité du grain est suffisamment grande, comme des attaques. Par la méthode du seuillage dynamique, ces trois attaques précédemment détectées sont sous le seuil, conséquence de la forte puissance locale du signal à ces endroits.

Lors de plusieurs tests, l'algorithme s'est montré assez robuste tant que l'on reste dans les conditions dictées au départ, qui étaient le non recouvrement des grains dans le temps. En effet, les attaques survenues dans un contexte de trop forte puissance locale, sont ignorées par le seuillage dynamique, situation qui arrive souvent dans un signal polyphonique. La méthode peut traiter des grains qui se recouvrent mais les résultats ne sont alors pas aussi fiables. Ainsi, en se restreignant dans ce contexte, nous avons donc grandement amélioré la méthode de détection d'attaque dans son application aux sons granulaires. Nous pouvons maintenant exploiter les indices temporels pour effectuer des analyses locales attribuables à chacun des grains sonores.

3.1.3 Extraction de paramètres

Une fois le signal analysé segmenté en grain, on peut par diverses méthodes d'analyse extraire les paramètres qui seront appliqués lors de la phase de resynthèse. Nous n'avons pas ici explorer toutes les possibilités d'analyse mais étions plutôt dans une approche de test et d'experimentation sur la validité de la technique d'un point de vue global. Nous avons pour valider la méthode procéder à l'étude d'un cas simple. Il s'agit de la pièce "Turenas" pour bande composée par John Chowning en 1977, entièrement synthétisée par techniques de synthèse FM et granulaire. Le principal intérêt de cette étude est que nous avons pour cette oeuvre l'entière partition qui a servi à la générer. Elle est sous la forme de fichier MUSICIV le programme de synthèse, dans la lignée des logiciels MUSIC n initialement créés par Max Mathews. John Chowning a beaucoup utilisé de distributions aléatoires pour les paramètres de synthèse, distributions que l'on pourra confronter avec les résultats de la méthode d'extraction.

Le spectrogramme de l'extrait que nous avons choisi d'analyser est représenté en figure 7. On peut très bien voir qu'il est composé de quatre parties, au sein desquelles il utilise différentes distributions aléatoires pour les paramètres de *pitch*. Pour confronter les résultats avec la partition, nous avons donc estimé la hauteur de chaque segment précédemment détecté.

Estimation du pitch Pour l'extraction du pitch de chaque grain nous avons utilisé le logiciel SuperVP, moteur de Audiosculpt, développé à l'IR-CAM par l'équipe Analyse-Synthèse. Ce programme est une boîte à outils

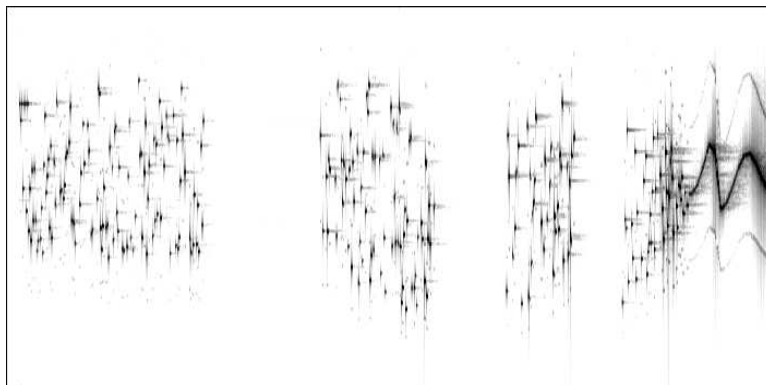


FIG. 7 – Spectrogramme pris entre 0 et 10000 Hz de l'extrait de Turenas de John Chowning.

permettant d'effectuer, sur le son, de nombreuses analyses et traitements avancés dans le domaine spectral. Il implémente notamment les techniques d'analyse par transformée de Fourier, par "Linear Predictive Coding", par méthode cepstrale, ... Il permet ainsi, et c'est ce qui va ici nous intéresser, de mettre en oeuvre des méthodes d'estimation de fréquence fondamentale. Nous avons utilisé ici l'analyse *pic* qui détecte les pics de plus forte énergie dans le spectre du signal analysé. On retient alors seulement le pic le plus haut qui nous donnera la hauteur estimée. Dans notre cas, cette technique offre des résultats fiables en grande partie du à la simplicité dans la structure spectrale des grains et au fait qu'ils ne se recouvrent pas dans le temps. Elle peut être toutefois inadéquate pour des sons polyphoniques ou dans le cas de fondamentale absente. On pourra alors préférer des analyses type *mask* (nous n'avons pas pu utilisé ces analyses, n'étant pas complètement implémentées dans la version de SuperVP que nous avons à disposition). Un des avantages de l'utilisation de ce logiciel réside dans la possibilité de n'effectuer les analyses qu'à des temps particuliers du son d'origine. Nous avons pu donc directement exploiter les temps d'attaques résultat de la méthode de segmentation. Voici la commande SuperVp utilisé pour l'analyse de notre extrait sonore :

```
supervp -v -t -Sseq4.aif -Apic 40.0 n1 -U -Oa :freq -N2048 -M2048
-Whanning -Ipos seq4.on seq4_picana.freq
```

Le résultat de cette analyse est présenté en figure 8. Nous n'avons finalement que peu de moyens de vérifier de manière précise ces résultats, ces valeurs n'étant déterminées dans la partition uniquement sous la forme

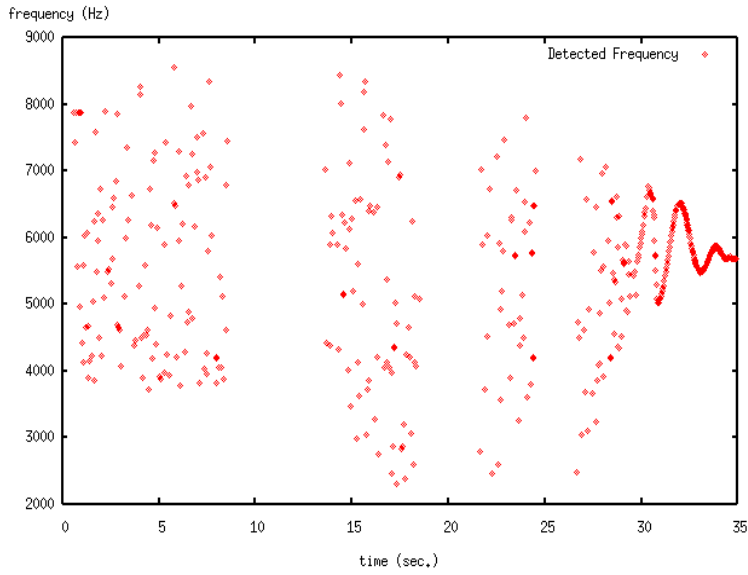


FIG. 8 – Resultat de l’estimation du pitch par analyse *pic*.

d’une distribution équiprobable entre deux bornes. Toutefois en superposant le spectre avec la figure 8, on se rend compte de la fiabilité des résultats de l’analyse.

Reste à confronter ceux ci avec les spécifications des paramètres décrites dans la partition *MUSICIV*. Le *pitch* est choisie par une fonction aléatoire équiprobable entre deux valeurs bornes. Ces bornes vont évoluer au sein des quatres parties de l’extrait suivant :

Temps	Borne Inf	Borne Sup
0 :89	E6	E7
8 :00	E6	E7
13 :00	E6	E7
18 :00	E5	E7
21 :00	E5	E7
24 :00	E6	E7
26 :00	E5	E7
34 :00	B6	B6

Nous avons donc maintenant à étudier les propriétés statistiques des résultats de l’extraction pour tenter de retrouver les données ayant servi à les générer. Une des première méthode qui vient à l’esprit est de calculer la moyenne et l’écart-type sur des fenêtres temporelles de taille à définir. Ceci

permettrait d’avoir dans ce cas des renseignements intéressants sur l’évolution des paramètres. Mais ce traitement n’est pas assez général, admettant de manière sous-jacente la propriété équiprobable de la distribution statistique. On souhaite en effet pouvoir extraire des distributions de probabilités arbitraires permettant la représentation de phénomènes bien plus variés. L’idée est alors de calculer les histogrammes des hauteurs extraites. Les paramètres évoluant relativement rapidement (passage de E5 à E6 de la borne inférieure en 3 secondes), nous devons calculer ces histogrammes sur des fenêtres temporelles assez courtes. Nous avons pris pour l’analyse de l’extrait des fenêtre de 1 seconde. Les histogrammes de la troisième partie sont représentés en figure 9.

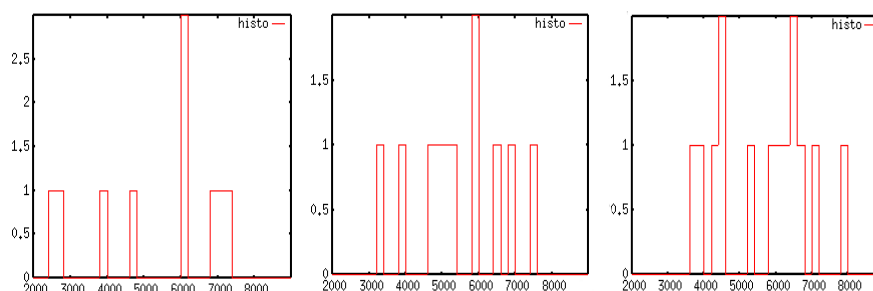


FIG. 9 – Histogrammes des fréquences correspondant à la troisième partie de l’extrait.

Ormis la tendance remarquable d’augmentation de la borne inférieure, on peut voir que ces résultats ne sont pas réellement significatifs par rapport à l’équiprobabilité de la distribution de départ. Ceci est dû à la faible quantité de donnée contenue dans le signal analysé. En effet, la densité des grains n’excède pas 15 grains/seconde dans cette partie et se pose ainsi le problème de la “rareté” des données pour l’extraction des distributions. Ce cas est tout de même particulier dans le sens où les paramètres régissant l’aléatoire évolue relativement vite dans le temps, ceci nous restreignant au calcul de l’histogramme sur une courte période. Ainsi, le calcul de l’histogramme a d’autant plus de sens que l’on se place sous l’hypothèse de lois stochastiques évoluant lentement au cours du temps. Dans ce cas, ce dernier pourra être calculé sur un nombre bien plus conséquent de donnée améliorant ainsi l’estimation de ces lois. Ici, une alternative à l’exploitation de l’histogramme serait de revenir au modèle plus simple de loi équiprobable bornée, ceci impliquant le seul calcul des valeurs minimales et maximales sur des courtes fenêtres temporelles.

3.1.4 Critique

Nous proposons ainsi une adaptation des techniques de segmentation de signaux audio à l'analyse granulaire permettant en définitive l'extraction de paramètre propre au modèle microsonore. L'analyse d'un cas simple nous a montré que sous certaines conditions la méthode donnait des résultats exploitables intéressants. L'extraction des paramètres doit toutefois être adaptée selon les propriétés des grains à détecter et leur modalité d'application en synthèse. En effet, ici les grains étant de type sinusoidaux, l'extraction de la fréquence fondamentale était donc justifiée, ce paramètre pouvant dès lors contrôler le *pitch* du synthétiseur granulaire. Mais, prenons le cas de grains moins élémentaires d'un point de vue spectral, comme contenant un bruit coloré. Il faudrait alors plutôt s'attacher à extraire des paramètres comme le centroïde spectral, plus apte à détecter des variations de hauteurs dans ce cas là. D'autre part, on peut vite se rendre compte de la difficulté de séparer les paramètres d'enveloppes des paramètres de la forme d'onde. Ainsi, l'étalement spectral pourra aussi bien être représenté par une forme d'onde complexe que par la longueur de l'enveloppe modulant une forme d'onde simple. On doit donc faire des choix sur les phénomènes extraits qui incombent à tel ou tel paramètre de synthèse, ceci en évitant autant que possible les relations redondantes.

Dans son application aux sons bruités, cette méthode représente une première approche intéressante sur les possibilités d'analyse. Elle peut donner des résultats exploitables pour les sons types craquements, impliquant des grains ayant une attaque franche. Mais, on ne peut toutefois que déplorer le caractère trop peu général de ce type d'analyse. La simple restriction de la détection des grains basée sur la dynamique du signal nous prive de résultats fiables sur de nombreuses classes de sons bruités. D'une part, une trop forte densité de grains, même de type "percussif", ne pourra être analysée de manière concluante, les fréquences de déclenchement des grains interférant avec leur contenu spectral. D'autre part, on ignore aussi tous les sons dont les grains ont des évolutions de dynamiques lentes, comme par exemple les sons de frottements ou de glissements. La seconde restriction principale est bien sûr le non-recouvrement temporel des grains. De nombreux type de sons bruités sont à priori composés de grains (ayant une structure commune) survenant en même temps ou ayant une longueur suffisante pour se recouvrir. Le bruit de l'eau qui coule est un exemple flagrant de ce type de sons. En fait, dès lors que sont en jeu dans le monde réel plusieurs entités émettrices (les feuilles d'un arbre), le son résultant (le vent dans les feuilles) est forcément un empilement temporel de grains. Avec cette méthode, l'extraction des paramètres devient dans ce cas vite problématique, n'ayant plus une relation

directe entre le phénomène extrait et le grain auquel il correspond. Ceci nous a donc poussé à rechercher d'autres voies possibles pour l'analyse permettant de passer au delà de ces limitations trop contraignantes. Comment décomposer alors un son en microentités dont les paramètres pourraient s'adapter à la synthèse granulaire ? On ne peut dès lors plus se baser sur des techniques de segmentation pour les raisons évoquées plus haut. L'idée est alors d'utiliser la projection orthogonale du signal analysé sur un ensemble de grains préalablement générés et c'est ce que nous allons évaluer maintenant dans notre étude de l'algorithme Matching Pursuit.

3.2 Analyse Matching Pursuit

3.2.1 Motivations

Suite au constat des limitations de l'analyse par segmentation temporelle, il nous fallait trouver une méthode qui permette une analyse bien plus flexible par rapport au modèle de signal requis. Notre but est effectivement de décomposer le signal analysé en entités sonore élémentaires. Repartons du modèle de départ selon lequel le signal analysé $x[n]$ peut être représenté comme une combinaison linéaire de fonctions élémentaires g_k selon :

$$x[n] = \sum_{k=1}^K \alpha_k g_k[n]$$

On voudrait représenter par ce modèle une grande variété de signaux de la manière la plus compacte possible, c'est à dire pouvoir représenter un signal par le jeu de coefficients α_k et fonctions g_k avec l'ensemble $\{i \in [1..K], \alpha_i > 0\}$ le plus petit possible. Il existe de nombreuses méthodes d'analyses se basant sur ce modèle. Les plus connues et usitées sont la transformée de Fourier et la transformée en ondelettes, toutes deux employant des fonctions g_k orthogonales entre elles. L'ensemble des g_k constitue ainsi dans ces cas une base sur laquelle est projeté le signal $x[n]$. Ces techniques ont beaucoup d'avantages notamment par le fait qu'elles proposent une décomposition $\{\alpha_k\}$ unique sur un ensemble réduit de fonctions $\{g_k\}$. Ceci est au détriment de la capacité de représentation d'une grande variété de signaux. En effet, avec la transformée de Fourier, un signal localisé en fréquence mais peu localisé en temps aura une représentation compacte. A l'inverse un signal localisé en temps (une impulsion) aura une représentation non-compacte qui caractérisera de manière beaucoup moins efficace la forme pourtant simple de départ. L'idée est alors, plutôt que d'avoir un ensemble de fonctions g_k orthogonales, de décomposer le signal analysé sur un ensemble redondant de fonctions paramétriques. Les paramètres rentrant en jeu doivent être en concordance avec ceux de la synthèse granulaire, ceci permettant de créer en discrétisant l'espace des paramètres un dictionnaire de grains servant de base à la décomposition. Plusieurs méthodes ont été proposées pour la décomposition sur de tels dictionnaires redondants. On pourra citer "Method of Frames" ou "Basis Pursuit", toutes deux proposant la résolution optimale globale du problème de décomposition. Nous allons nous intéresser ici particulièrement à l'algorithme "Matching Pursuit" initialement proposé par S. Mallat et Z. Zhang dans [11], qui comme les deux autres méthodes estime les α_k mais de manière itérative en affinant sa solution à chaque pas. Cela permet d'avoir un contrôle flexible du nombre de grains avec lequel on

souhaite représenter le signal, notamment par la possibilité de reprise d’une décomposition trop grossière.

L’intérêt pour l’analyse granulaire est ici multiple. D’un part, on peut analyser une bien plus grande variété de signaux qu’avec l’analyse par segmentation, la structure de la méthode permettant la détection de grains se recouvrant dans le temps. De plus, l’extraction des paramètres suivant la décomposition est ici beaucoup plus directe. En effet, les fonctions sur lesquelles on projette le signal à analyser étant elles mêmes définies de façon paramétrique, il y correspondance directe entre le choix d’un grain et les paramètres applicables en synthèse. Toutefois il y a tout de même des limitations quand au choix des fonctions g_k . Cette méthode suppose en effet l’utilisation de fonctions déterministes simples dont les plus usités sont les gaborettes (sinusoides modulées par une gaussienne).

Nous allons maintenant decrire brièvement la méthode pour ensuite procéder à une étude de cas, nous permettant d’évaluer l’intérêt de celle ci, dans son application à l’analyse granulaire.

3.2.2 Décomposition Matching Pursuit

Matching Pursuit est un algorithme itératif “glouton” (le nombre d’itération est directement lié à la finesse de l’approximation) qui effectue une décomposition de signaux sur une base d’atomes choisis dans un dictionnaire. Ce dernier a la propriété d’être redondant, c’est à dire, plus que de proposer une base de fonctions orthogonales, il inclue une grande variété de structures fréquentielles et temporelles permettant la représentation compacte de nombreux comportements différents. A chaque itération, l’atome qui approxime le mieux le signal est choisi ; ensuite la contribution de cet atome au signal est soustraite et on applique le même schéma sur le résidu ainsi calculé. En utilisant la norme-2 comme mesure d’approximation, la tâche de l’algorithme à la i -ème itération est de trouver l’atome $g_{m(i)}[n]$ qui minimise la norme-2 du signal résiduel :

$$r_{i+1}[n] = r_i[n] - \alpha_i g_{m(i)}[n]$$

avec α_i le coefficient correspondant au poids de la contribution de l’atome dans le signal et $m(i)$ l’index de l’atome dans le dictionnaire choisi à l’étape i . L’algorithme commence avec $r_1[n] = x[n]$ le signal original. En considérant les signaux comme des vecteurs colonnes, l’atome optimal que l’on doit choisir à l’étape i peut-être exprimé par :

$$g_{m(i)} = \arg \min_{g_{m(i)} \in D} \|r_{i+1}\|^2 = \arg \min_{g_{m(i)} \in D} \|r_i - \alpha_i g_{m(i)}\|^2$$

avec D le dictionnaire d’atomes. Le principe d’orthogonalité nous donne la valeur de α_i :

$$\begin{aligned}\langle r_{i+1}, g_{m(i)} \rangle &= \langle r_i - \alpha_i g_{m(i)}, g_{m(i)} \rangle = (r_i - \alpha_i g_{m(i)})^H g_{m(i)} = 0 \\ \Rightarrow \alpha_i &= \frac{\langle g_{m(i)}, r_i \rangle}{\langle g_{m(i)}, g_{m(i)} \rangle} = \frac{\langle g_{m(i)}, r_i \rangle}{\|g_{m(i)}\|^2}\end{aligned}$$

la dernière étape impliquant que les atomes soit de norme unitaire. L'erreur d'approximation représentée par la norme du résidu r_{i+1} peut alors être exprimée par :

$$\|r_{i+1}\|^2 = \|r_i\|^2 - \frac{|\langle g_{m(i)}, r_i \rangle|^2}{\|g_{m(i)}\|^2} = \|r_i\|^2 - |\alpha_i|^2$$

qui est minimal en maximisant $|\alpha_i|^2 = |\langle g_{m(i)}, r_i \rangle|^2$. Cela revient donc, à chaque itération de l'algorithme, à choisir l'atome qui a le plus grand coefficient de corrélation $|\alpha_i|$ c'est dire :

$$g_{m(i)} = \arg \max_{g_{m(i)} \in D} |\langle g_{m(i)}, r_i \rangle|$$

L'expression de l'erreur d'approximation $\|r_{i+1}\|^2$ nous montre qu'elle décroît avec le nombre d'itération de l'algorithme. Nous avons donc plusieurs choix quand à la clause d'arrêt de ce dernier : soit de décider au préalable le nombre d'atomes avec lesquels on veut représenter notre signal de départ, soit de s'arrêter quand l'erreur d'approximation passe sous un certain seuil influençant la finesse de la décomposition. Après I itérations, l'algorithme donne l'estimation du signal :

$$x[n] \approx \hat{x}[n] = \sum_{i=1}^I \alpha_i g_{m(i)}[n]$$

Pour pouvoir représenter de larges classes de signaux différents, l'algorithme Matching Pursuit utilise des dictionnaires comportants une grande quantité d'atomes. A chaque étape, comme on l'a vu, on calcule la corrélation $\langle g, r_i \rangle$ pour chaque atome g ce qui est coûteux en temps de calcul. En fait, comme la montré [11] le calcul de la corrélation peut se faire de manière récursive en remarquant que :

$$\langle g, r_{i+1} \rangle = \langle g, r_i \rangle - \alpha_i \langle g, g_{m(i)} \rangle$$

Le seul terme à mettre à jour, à l'étape $i + 1$, est alors le terme $\langle g, g_{m(i)} \rangle$, corrélation entre chaque atome du dictionnaire, qui peut être pré-calculée.

On peut noter que la qualité de l'approximation est grandement dépendante des propriétés des atomes contenues dans le dictionnaire. L'intérêt est que ces atomes puissent décrire une bonne variété de phénomènes temporels et fréquentiels. On rejoint ici l'idée de quanta sonore imaginé par Gabor. Les atomes de Gabor (ou gaborettes) sont en effet particulièrement adaptés à une description temps-fréquence fine des signaux analysés.

Atomes de Gabor La littérature sur l'algorithme Matching Pursuit a largement abordé l'utilisation des atomes de Gabor pour la décomposition temps-fréquences de signaux. Ces propriétés temps-fréquences sont aisément contrôlable paramétriquement. Ces atomes sont construits en mettant à l'échelle, en modulant et en translatant une fonction gaussienne g de norme unitaire définie par :

$$g[n] = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{n^2}{2\sigma^2}\right)$$

Chaque triplet $\{s, \omega, \tau\}$, avec échelle $s > 0$, fréquence de modulation ω et translation τ , définissent un atome du dictionnaire suivant :

$$g_{\{s, \omega, \tau\}}[n] = \frac{1}{\sqrt{s}} g\left(\frac{n - \tau}{s}\right) e^{j\omega n}$$

A chaque triplet $\{s, \omega, \tau\}$ correspond des propriétés temps-fréquences bien spécifiques. Ainsi, la fonction $g_{\{s, \omega, \tau\}}[n]$ est centrée sur l'abscisse τ et son énergie est concentrée au voisinage de τ , dont la taille est proportionnelle à s . Dans le domaine spectral, sa transformée de Fourier est centrée sur la fréquence ω et son énergie est concentrée au voisinage de ω , dont la taille est proportionnelle à $1/s$. On pourra remarquer que la décomposition avec de tels atomes n'est qu'une généralisation des méthodes type Fourier ou Ondelettes. Mais on s'attache ici à ne pas restreindre l'espace des paramètres comme dans ces dernières, ceci étant la condition à une représentation adaptative des signaux analysés.

Cependant La fonction g étant symétrique, on ne peut à priori représenter localement que des phénomènes temporels symétriques. Ceci est problématique pour représenter des phénomènes asymétriques comme les *transients* qui sont fréquents dans les sons naturels. La représentations de tels phénomènes par atomes de Gabor conduit à un artefact de pré-écho, correspondant à un ajout erroné d'énergie avant le *transient*. Une solution à ce problème a été proposé par Jaggi, Gribonval al.dans [8] puis reprise par R. Gribonval dans [1] et [5], dans une extension de Matching Pursuit, nommée High resolution Matching Pursuit, dont l'idée a été de modifier la fonction de corrélation

pour pénaliser les atomes introduisant de tels artefacts. Une autre solution proposée par Goodwin et Vetterli dans [3] consiste à utiliser des atomes asymétriques telles des sinusoides amorties.

Atomes sinusoidaux amortis La décomposition de signaux par sinusoides amorties permet de représenter des phénomènes temporels asymétriques sans artefact de pré-écho. De plus, l'utilisation de ce type de structure pour l'analyse de sons naturels peut se justifier par son caractère très courant du point de vue physique. En effet, incarnant la propriété essentielle de causalité, de nombreuses mesures de phénomènes naturels suivent de telles lois, comme par exemple le mouvement d'un ressort ou les sons percussifs d'un xylophone.

Comme les atomes de Gabor, les atomes sinusoidaux amorties peuvent être indexés dans le dictionnaire par leurs paramètres caractéristiques : le facteur d'amortissement $a < 1$, la fréquence de modulation ω et la translation τ définissent ces atomes :

$$g_{\{a,\omega,\tau\}}[n] = S a^{(n-\tau)} e^{j\omega n} u[n - \tau]$$

ou S représente le facteur d'échelle pour satisfaire la normalisation de l'atome et u fonction rectangulaire permettant de tronquer l'atome lorsque a^i est en dessous d'un certain seuil T . On pourra noter que l'on utilise ici le paramètre de facteur d'amortissement a et non le facteur d'échelle s des atomes de Gabor, le jeu de paramètres indexant le dictionnaire devenant $\{a, \omega, \tau\}$. On pourra toutefois lui faire correspondre une taille $L = \lceil \log T / \log a \rceil$.

Autres Atomes L'utilisation d'autres types d'atomes a aussi été explorée. Ainsi, R. Gribonval propose dans [4] l'utilisation de *chirplet* pour la décomposition. Il s'agit d'atomes de Gabor mais dont la fréquence modulante varie linéairement. En effet il sont caractérisés par les mêmes paramètres plus le paramètre de *chirp rate* qui détermine la vitesse d'évolution de la fréquence. Bien que ce type de grains peut être intéressant pour représenter des variations fréquentielles continues, nous nous y sommes pas directement intéressés car leur application en synthèse n'était pas réalisable à court terme.

En réalité il existe de nombreuses variantes des grains de Gabor qui restent applicables pour l'algorithme Matching Pursuit. Ils ne diffèrent de ces derniers que par la forme de la fonction modulée (de l'enveloppe). Ainsi, on peut par exemple évoquer les grains FoF, pour Forme d'onde Formantique, permettant une décomposition plus adaptée au signaux vocaux.

3.2.3 Etude de cas

Nous proposons pour évaluer l'application de cette méthode à l'analyse granulaire l'étude d'un exemple de decomposition. Nous avons pour cela utilisé le logiciel *LastWave* développé par R. Gribonval qui implémente de manière assez complète l'analyse Matching Pursuit. Il s'agit en fait d'un interpréteur scientifique, dans la lignée de logiciel type *Matlab*, permettant d'effectuer des scripts de calculs relativement facilement et d'ensuite visualiser les résultats de manière graphique. Bien que cela ait impliqué le test des seules extensions implémentés de l'algorithme, cela nous donne tout de même une bonne appréhension de la faisabilité de son utilisation en analyse granulaire.

Sons de ruisseau On est parti du postulat que les sons d'eau pouvaient être représentés par des grains simples du point de vue de leur structure. En effet, d'un point de vue perceptif, on peut remarquer que ceux-ci ont des composantes sinusoidales relativement marquées au niveau local. On ne peut toutefois pas y attribuer de hauteurs précises, mais on peut sentir une evolution très rapide de composantes frequentiellles simples, ceci motivant son analyse par Matching Pursuit.

Le spectrogramme du son d'eau analysé est représenté en figure 10.

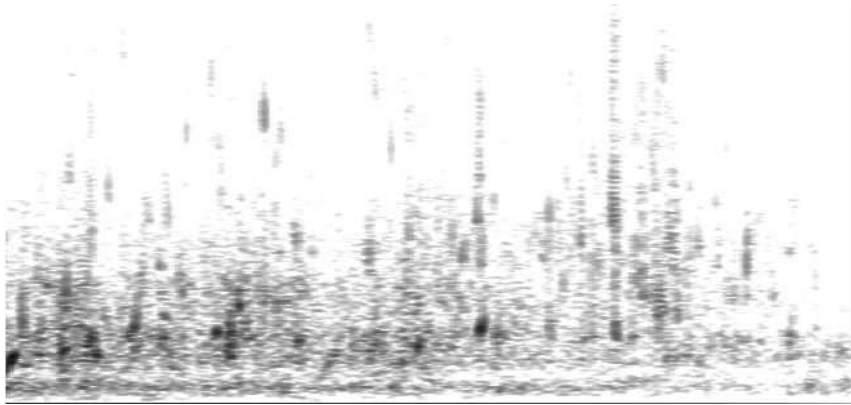


FIG. 10 – Spectrogramme du son d'eau analysé d'une durée de 1.8 secondes pris entre 0 et 10000 Hz

Il nous faut déterminer maintenant avec quels types de grains on peut le mieux représenter ce signal. Nous avons restreint notre choix seulement entre atomes de Gabor et sinusoides amorties étant les représentants les plus particuliers des structures sonores locales à extraire. On peut utiliser comme

critère de choix la comparaison de l'erreur d'approximation $\|r_{i+1}\|^2$ en fonction du nombre d'itération de l'algorithme. Nous avons donc effectué une décomposition sur le nombre important de 3000 itérations. Les erreurs respectives d'approximation sont représentées en figure 11.

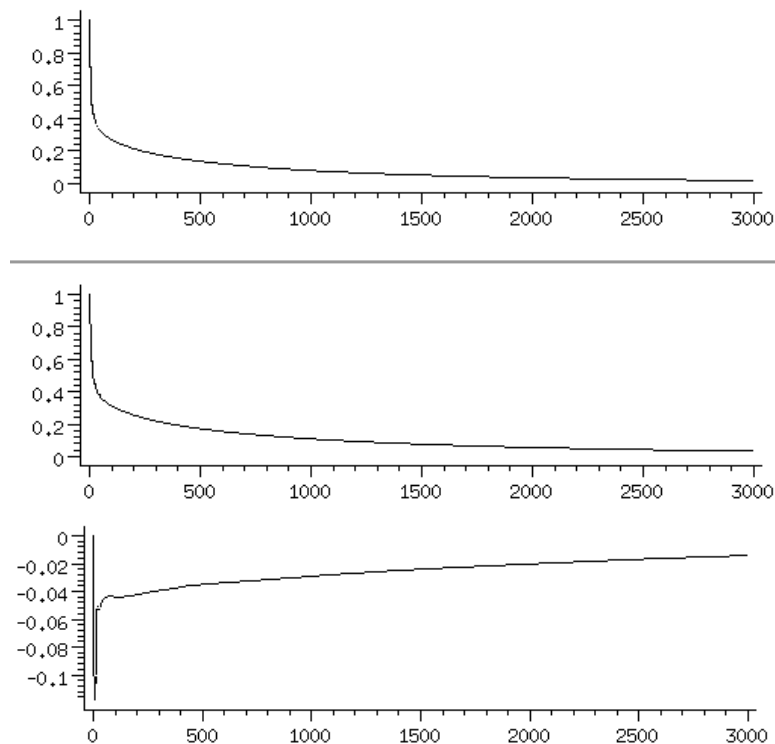


FIG. 11 – Erreurs d'approximation en fonction du nombre d'iterations respectivement pour les atomes de Gabor et pour les atomes sinusoides amorties. Le troisième graphe représente la différence de l'erreur Gabor à l'erreur sinusoides amortie

On peut remarquer que la pente des courbes d'erreurs, forte aux premières itérations (jusqu'à environ 50 iterations), a tendance à s'adoucir par la suite. Ceci est du à la détection en premier lieu de long atomes correspondant aux composantes fréquentielles perceptivement discernables et représentant une grande part de l'énergie totale du signal. Concernant le choix du type d'atome, les courbes ont globalement la même allure mais celle des atomes sinusoidaux amorties décroissent moins vite. En effet, à la 3000^{ième} itérations l'erreur est de 0.034 contre 0.019 pour les gaborettes. Nous avons donc choisi d'effectuer la decomposition sur les atomes de Gabor.

Nous devons donc déterminer maintenant le nombre d'atomes nécessaires

pour une bonne représentation de notre signal. Cela revient à savoir quelle erreur d'approximation somme nous prêt à accepter. Ce choix doit se faire au regard de l'évolution de cette erreur qui soustend l'intérêt ou non d'une poursuite de l'algorithme. En effet, on peut s'arrêter quand une itération supplémentaire n'apporte plus d'information assez significative sur le contenu du signal. Sur la courbe d'erreur, on peut remarquer qu'à partir de 2000 atomes l'erreur a une valeur suffisamment basse de 0.05 et n'a plus vraiment d'évolution significative. Ceci correspond par ailleurs à une densité relativement élevée de 1000 grains/secondes au niveau de la resynthèse.

Le résultat de la décomposition est représenté sur la figure 12. Il s'agit d'une visualisation temps-fréquence proposée par *LastWave*. Elle représente chaque grain sous forme d'un "blob" en deux dimension correspondant à son énergie théorique dans l'espace temps-fréquence. Les traits verticaux correspondent à des atomes très courts donc peu localisés en fréquence. A l'inverse les traits horizontaux représentent des atomes longs peu localisés en temps.

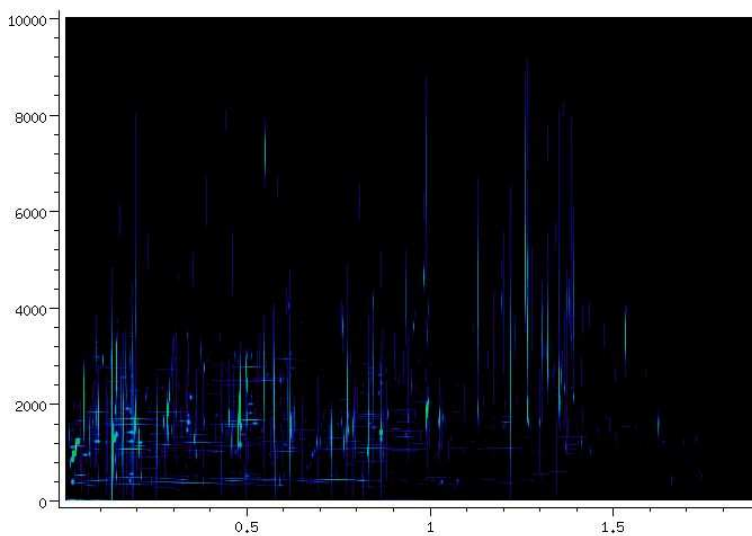


FIG. 12 – Résultat de la décomposition Matching Pursuit du bruit d'eau entre 0 et 10000 Hz

On peut voir en confrontant le spectrogramme avec le resultat d'analyse que les structures du son de départ semblent bien représentées. Notamment, les grains isolés repérables au dessus de $5000Hz$ ont bien été détectés. De plus, l'analyse met à jour de manière notable la faible baisse de la puissance du signal au cours du temps. Ainsi, la décomposition du son en atomes de Gabor ainsi effectué, on peut maintenant exploiter les données des atomes

déTECTÉS pour les appliquer à la synthèse.

De la même manière que pour l'analyse par segmentation, on va ici calculer les histogrammes des différents paramètres définissant les grains du dictionnaire. Ici, ces paramètres, que sont l'amplitude, la fréquence centrale, la durée et le temps d'occurrence, sont directement applicables à la synthèse. De par le caractère relativement constant du son de départ, nous avons choisi d'effectuer les histogrammes sur toute la durée du son. Nous étions, par ailleurs, plus soucieux d'extraire les propriétés de la classe des sons d'eau plutôt que de coller au plus près de ce représentant. Les histogrammes d'amplitude, fréquence et durée des atomes sont représentés en figure 13.

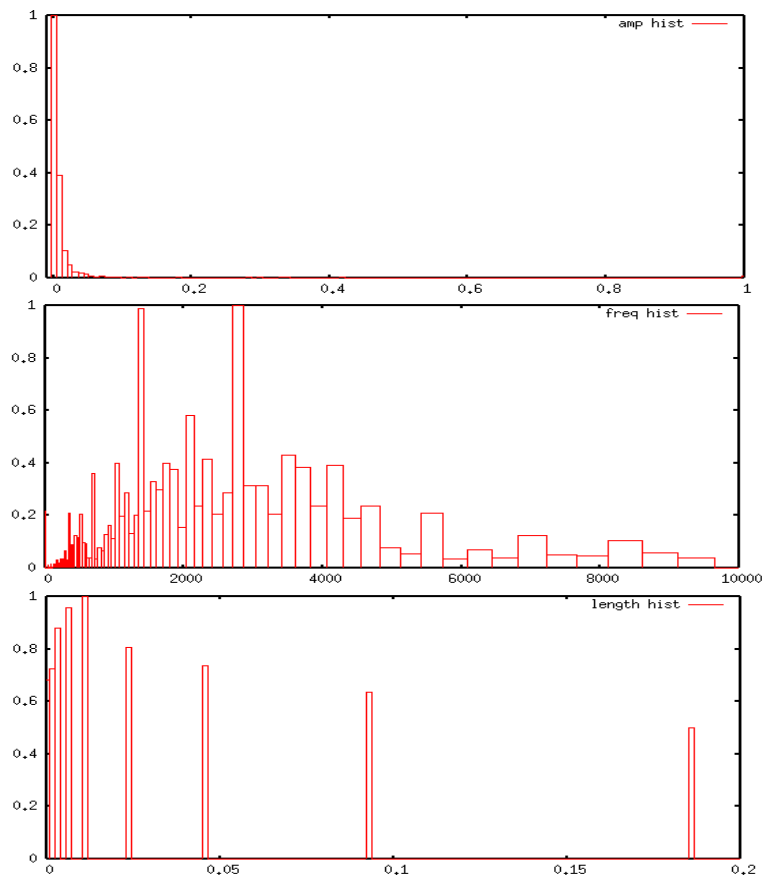


FIG. 13 – Histogrammes normalisés des paramètres des grains détectés, respectivement d'amplitude, de fréquence en Hz et de durée en seconde.

Nous nous devons d'émettre quelques commentaires sur ces graphes. Tout d'abord, nous avons calculé ces histogrammes sur 127 valeurs pour les rendre

compatibles avec le générateur aléatoire du module de synthèse. De plus, bien que l'échelle des fréquences des atomes du dictionnaire soit linéaire, nous avons préféré, au regard des considérations psychoacoustiques sur la discrimination des variations fréquentielles en fonction de la hauteur, d'utiliser une échelle logarithmique pour l'histogramme. Ceci était de plus en accord avec le module de synthèse granulaire qui acceptait des distributions stochastiques sous cette forme. Par ailleurs à propos de l'histogramme d'amplitude, on peut remarquer que les valeurs sont très ramassées sur 0, avec un maximum de nombre d'occurrence à 0. Il ne s'agit bien évidemment pas de la réelle amplitude détectée mais a à voir avec la discrétisation impliquée par l'utilisation de l'histogramme. Pour contourner ce problème, on peut diminuer le pas de l'histogramme en prenant bien en compte la relation qu'il maintient avec le nombre de données analysées. Un autre point remarquable est la forme de l'histogramme des durées des atomes. Elle est directement liée à la discrétisation des paramètres lors de la construction du dictionnaire. En effet, le logiciel *LastWave* utilise uniquement des longueurs d'atomes sous la forme 2^p . Ce choix a apparemment été fait pour optimiser le calcul de la corrélation entre les atomes. Cela peut dans notre cas être problématique ayant une résolution de plus en plus faible pour les longues durées. Néanmoins, nous voulons ici évaluer la méthode Matching Pursuit et non le logiciel lui-même. Il nous faut simplement considérer cette limitation dans nos résultats, la méthode générale n'ayant pas de telles restrictions.

Il a aussi fallu pour appliquer ces résultats en synthèse estimer la fréquence d'occurrence des grains. Or les temps d'occurrence sont eux aussi discrétisés amenant ainsi à la détection de nombreux atomes synchrones (au même temps). Or le module de synthèse ne pouvait déclencher deux grains exactement synchrones. C'est pourquoi, en remarquant que le nombre de grains détectés par unité de temps était relativement constant, nous avons choisi d'utiliser, pour déterminer la fréquence de déclenchement, la moyenne des différences de temps de détection entre deux atomes successifs. Nous avons donc approximé cette fréquence à $F = 1/0.0009 \simeq 1000Hz$.

Application en synthèse L'extraction des paramètres et leurs propriétés stochastiques ainsi faites, nous pouvons maintenant appliquer ces résultats en synthèse pour évaluer la validité de l'analyse. Nous avons donc en premier lieu, en accord avec les choix d'analyses, utilisé des grains d'enveloppe gaussienne modulée par une forme d'onde sinusoidale. Mais le son résultant, bien que pouvant se rapprocher d'un bruit d'eau dans sa distribution fréquentielle, n'était pas réellement probant d'un point de vue perceptif. Les composantes fréquentielles discernables dans le son naturel étaient en effet trop marquées,

rendant le son de connotation trop synthétique. Nous avons alors expérimenté l'utilisation d'une autre enveloppe de grains et nous avons remarqué que le choix d'une enveloppe à deux segments exponentiels donnait des meilleurs résultats. En effet, le sons d'eau était plus bruiteux et les trop fortes composantes fréquentielles de l'enveloppe gaussienne était alors, bien que perceptibles, beaucoup plus naturelles. Le spectre de ce son resynthétisé est représenté en figure 14.

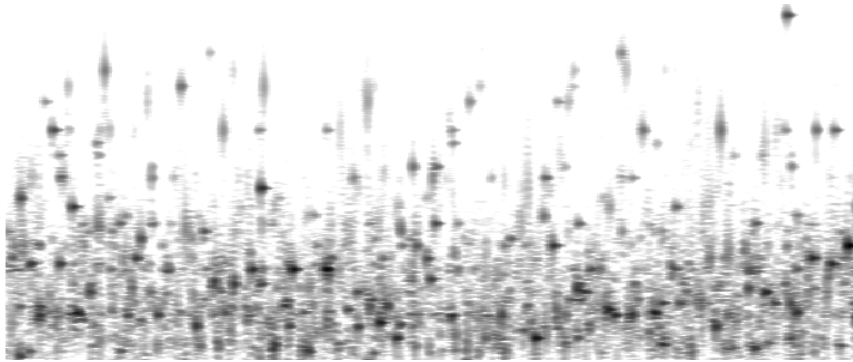


FIG. 14 – Spectre du son d'eau resynthétisé avec enveloppe à segment exponentiel .

Les résultats décevants de la resynthèse par enveloppe gaussienne peuvent s'expliquer par le fait d'ignorer la corrélation qui peut exister entre les différents paramètres. Nous n'avons en effet pas pris en compte les relations qui pouvaient lier les paramètres entres eux. Ceci implique au niveau de la synthèse une architecture beaucoup plus complexe dont la réalisation ne pouvait pas rentrer dans le cadre du stage. Pourtant, il semble que cela soit un phénomène important pour l'analyse et la resynthèse de son bruités ayant à priori une grande densité granulaire. Dans notre cas, les composantes fréquentielles principales du son naturel peuvent être synthétisées avec une enveloppe plus grande et/ou une amplitude plus forte que celles avec lesquels elles étaient détectés. Cela est d'autant plus flagrant que l'on a extrait une grand quantité de grains à des amplitudes variées, les paramètres des plus faibles influençant ainsi ceux des plus forts. Cet artefact peut ainsi être atténué avec le choix d'une enveloppe moins puissante et perceptiblement moins longue comme c'est le cas de l'enveloppe à deux segments exponentiels.

Critique Les résultats sont encourageants. En effet, cette correction ainsi faites, le son resynthétisé est bien attribuable à un son d'eau, bien qu'il faille des expériences perceptives dans les règles pour s'en assurer. Ainsi, la

méthode Matching Pursuit est intéressante pour l'extraction des propriétés granulaires d'un son naturel. Nous devons toutefois nous restreindre aux seuls sons dont la structure peut à priori être représentée par une somme de grains simples. Les limitations et choix d'optimisations du logiciel *LastWave*, comme par exemple la discrétisation du paramètre de longueur, n'étant pas relative à la méthode, cela permet d'envisager sa réimplémentation permettant une configuration plus poussée, notamment dans la construction du dictionnaire. En effet, il s'agit du point réellement le plus sensible de la méthode, et nous nous devons de pouvoir l'adapter pour orienter son application à la synthèse granulaire. Toutefois, cette algorithmique ne peut à priori pas s'appliquer à tout type de son bruités. Les sons composés de grains de structure plus complexe peuvent être décomposés par celui-ci mais faisant l'impasse sur la "cohérence granulaire" représentant une spécification forte dans notre démarche. Un bon contre exemple pourrait être les sons de grattements dont les grains ne peuvent être considérés comme sinusoidaux. Il a donc fallu ainsi rechercher une autre approche permettant une décomposition granulaire plus souple et c'est ce que nous allons détailler maintenant dans notre proposition d'extension de Matching Pursuit au domaine spectral.

3.3 Analyse Spectral Matching Pursuit

3.3.1 Motivations

Nous venons de voir les possibilités très intéressantes qu'offrent l'algorithme Matching Pursuit pour la décomposition de signaux suivant un ensemble de grains élémentaires. Un nombre assez important de classes de sons bruités peut à priori être représenté par de tels atomes. Mais il ne faut pas perdre de vue la cohérence granulaire suivant laquelle on souhaite décomposer les sons analysés. Une représentation Matching Pursuit, aussi fidèle soit elle, ne sera pas nécessairement satisfaisante dans le cas général. Les sons de grattement en sont un bon contre-exemple. L'algorithme pourra trouver une approximation fidèle mais n'en conservera pas la propriété granulaire intrinsèque, éclatant le grain complexe qui le caractérise en ensemble de sous-grains sinusoidaux. Ceci nous amène à revoir la méthode dans le but de l'appliquer à la décomposition des sons en grains arbitraires. Nous avons souhaité garder le principe de base de l'algorithme Matching Pursuit car il est bien adapté au modèle de son granulaire par l'extraction des propriétés locales d'un signal. L'emploi de dictionnaires offre notamment de sérieux avantages quand à l'application des résultats pour la synthèse.

Nous avons donc la tâche d'adapter cet algorithme à la décomposition sur un ensemble de grains arbitraires. Une première idée serait de construire un nouveau dictionnaire à partir des grains avec lesquels on veut décomposer le signal. Mais un problème survient dès lors que les grains ne sont plus déterministes (en tout cas entièrement). En effet, prenons un exemple simple pour argumenter ce propos. Supposons que l'on souhaite décomposer le signal sur un dictionnaire comportant un unique atome défini comme une séquence $G[k]$ de longueur K de variables IID (indépendantes identiquement distribuée) de loi gaussienne $\mathcal{N}(0, 1)$. On veut analyser un signal $X[k]$ de longueur K constitué d'une suite de variable IID de même loi $\mathcal{N}(0, 1)$. L'algorithme cherche alors à calculer la corrélation $\langle X, G \rangle$ qui s'écrit :

$$\langle X, G \rangle = \sum_{i=0}^K X[i]G[i]$$

Cette grandeur est une variable aléatoire et son espérance s'écrit, en notant que les $X[i]$ et $G[i]$ sont indépendants :

$$E(\langle X, G \rangle) = \sum_{i=0}^K E(X[i]G[i]) = \sum_{i=0}^K E(X[i])E(G[i])$$

$$E(\langle X, G \rangle) = K\mu^2 \quad \text{avec} \quad \mu = E(X) = E(G) = 0$$

Ayant une espérance nulle pour la fonction de corrélation, l'algorithme n'est ainsi pas en mesure de détecter dans le signal le grain pourtant de structure identique à celui du dictionnaire. Cet exemple simple n'est pas réellement caractéristique d'analyse réelle mais il nous montre tout de même comment les composantes stochastiques des grains du dictionnaire peuvent fausser la détection. Il nous faut donc, si l'on souhaite garder la même structure pour l'algorithme, trouver une autre fonction de corrélation adaptée aux signaux non-déterministes. Une idée serait d'utiliser le produit scalaire de l'autocovariance empirique pour comparer le contenu spectral de l'atome g avec le signal x réduit au domaine de g suivant :

$$C(x, g) = \langle acov_x, acov_g \rangle \quad \text{avec} \quad acov_z(d) = \sum_{k=0}^{N-d} z(k) * z(k + d)$$

Mais plusieurs problèmes surviennent alors. D'une part on perd l'information de phase qui peut lors de la soustraction du grain au signal ajouter de l'énergie plutôt qu'en enlever. En effet, deux grains sinusoidaux g_1 et g_2 en opposition de phase ayant la même autocovariance, la soustraction de l'un par l'autre donnera $x = 2 * g_1$. De plus, l'autocovariance ne conserve pas non plus les évolutions temporelles dans la structure du grain. Deux grains $g_1(t)$ et son renversé temporel $g_2(t) = g_1(N - t)$ auront aussi la même autocovariance, l'algorithme perdant ainsi un pouvoir de discrimination essentiel entre différents grains complexes. Ces constats nous empêchent donc d'envisager l'utilisation de l'autocovariance comme fonction de corrélation. Il nous faut donc trouver une méthode permettant de comparer la structure à la fois temporelle et spectrale des grains avec celle du signal. Nous venons de voir les limites du travail dans le domaine temporel pour prendre en compte ces spécifications. Il y a bien des possibilités d'extension de la fonction de corrélation mais elles ne font qu'alourdir le principe de départ simple de l'algorithme.

L'idée est alors de garder le principe de décomposition adaptative de Matching Pursuit et de l'adapter au domaine spectral qui permet une représentation bien plus fine des propriétés structurelles de signaux variés. En effet, par des techniques de lissage de spectre, on peut représenter des signaux non déterministe donnant approximation de leur densité spectrale de puissance. Il ne s'agit plus de comparer le signal avec les atomes sous leur forme temporelle mais d'utiliser plutôt pour cela leur spectrogramme. L'avantage des spectrogrammes par rapport à une unique transformée de Fourier est de conserver des informations temporelles relativement précises, chose que ne peut pas faire la TF, comme nous l'avons vu plus haut. Cette idée pourrait s'apparenter à des techniques de reconnaissance de formes spectrales à partir d'un

dictionnaire. Nous allons maintenant détailler la démarche d'une telle analyse, que l'on a nommé (sans beaucoup d'imagination) Spectral Matching Pursuit.

3.3.2 L'algorithme Spectral Matching Pursuit

Il nous faut partir d'ici d'un autre modèle de son granulaire équivalent au premier introduit plus haut mais adapté au domaine spectral. Nous étions parti du postulat que le signal x de départ pouvait se décomposer en somme pondérée de K atomes g_k selon :

$$x[n] = \sum_{k=1}^K \alpha_k g_k[n]$$

En introduisant la transformée de Fourier à court terme TFCT $X(t, f)$ de x au temps t de longueur N définie par :

$$\begin{aligned} X(t, f) &= \sum_{n=1}^N w[n] x[t+n] e^{2j\pi f} \\ &= \sum_{n=1}^N w[n] \left(\sum_{k=1}^K \alpha_k g_k[t+n] \right) e^{2j\pi f} \\ &= \sum_{k=1}^K \alpha_k \left(\sum_{n=1}^N w[n] g_k[t+n] e^{2j\pi f} \right) \end{aligned}$$

Ainsi :

$$X(t, f) = \sum_{k=1}^K \alpha_k G_k(t, f)$$

Avec $G_k(t, f)$ TFCT de g_k au temps t de taille N . En toute rigueur, il faudrait travailler avec le spectre complexe $X(t, f)$ mais ceci impliquait plusieurs contraintes qui pouvaient poser problème dans la mise en oeuvre de l'algorithme. Le but est ici de projeter le signal $X(t, f)$ sur les atomes $G(t, f)$ du dictionnaire. Lors de la construction du dictionnaire, ce dernier étant fini, chacun des paramètres, dont la translation temporelle de l'atome, est ainsi discrétisé selon les besoins de la décomposition. Or étant peu envisageable, d'un point de vue temps de calcul, d'effectuer le calcul de la corrélation pour chacun des atomes à chaque échantillon, le pas temporel n'est donc pas une donnée négligeable. Il introduit ainsi des problèmes de déphasage qui peuvent résulter en une pénalisation du choix d'atomes pourtant représentatif

du signal analysé. En réalité, l'information de phase du spectre n'est pas vraiment primordiale dans notre cas s'attachant plus à détecter des formes spectrales associables à un jeu de paramètres. Nous avons donc choisi de travailler exclusivement sur le module du spectre $|X(t, f)|$. Nous supposons donc que $|X(t, f)|$ peut être décomposé selon :

$$|X(t, f)| = \sum_{k=1}^K \alpha_k |G_k(t, f)|$$

Il est important de voir que l'on travaille maintenant avec des matrices et non plus des vecteurs temporels. Pour simplifier les notations et la structure de la méthodes, nous avons fait en sorte de manipuler le module des spectrogrammes sous la forme de vecteurs $X[p_{t,f}]$:

$$X[p_{t,f}] = \left| \sum_{n=0}^N w[n] x[t * d + n] e^{2j\pi f n} \right|$$

$$\text{avec l'indice } p_{t,f} = tN + f/F_e$$

avec N taille de la TFCT et d pas d'avancement pour le calcul des TFCT réglant notamment leur recouvrement.

Comme précédemment, on veut donc à chaque étape i de l'algorithme trouver dans le dictionnaire D l'atome $G_{m(i)}$ qui, par projection du résidu R_i minimise la norme-2 de R_{i+1} défini par :

$$R_{i+1} = R_i - \alpha_i G_{m(i)}$$

avec $m(i)$ designant l'index de l'atome choisi dans le dictionnaire à l'étape i . L'algorithme commence avec $R_1 = X$. Le but est de trouver $G_{m(i)}$ selon :

$$G_{m(i)} = \arg \min_{G_{m(i)} \in D} \|R_{i+1}\|^2$$

Par ailleurs :

$$\langle R_{i+1}, G_{m(i)} \rangle = \langle R_i, G_{m(i)} \rangle - \alpha_i \langle G_{m(i)}, G_{m(i)} \rangle$$

Par le théorème de la projection orthogonale on a :

$$\langle R_{i+1}, G_{m(i)} \rangle = \langle R_i - \alpha_i G_{m(i)}, G_{m(i)} \rangle = 0$$

D'où :

$$\alpha_i = \frac{\langle R_i, G_{m(i)} \rangle}{\langle G_{m(i)}, G_{m(i)} \rangle} = \langle R_i, G_{m(i)} \rangle$$

impliquant ainsi que la norme-2 des atomes $\|G_k\| = 1$. On peut alors en deduire la norme-2 du signal residuel $\|R_{i+1}\|$:

$$\|R_{i+1}\|^2 = \|R_i\|^2 - |\alpha_i|^2$$

Minimiser $\|R_{i+1}\|^2$ revient donc à maximiser $|\alpha_i|^2$. On doit donc choisir à chaque itération i l'atome qui aura le plus grand coefficient de corrélation $\alpha_i = \langle R_i, G_{m(i)} \rangle$.

estimation amplitude Une fois trouvés tous les coefficients α_i correspondant à la décomposition, il s'agit d'en tirer les paramètres d'amplitude applicables à chacun des grains dans le domaine temporel. En effet, il n'y a pas d'égalité directe entre corrélation et amplitude de part la normalisation des atomes. Soit $x = \gamma g$, on a alors le module du spectre $X = \gamma G$. Le calcul de la corrélation nous donne $X = \alpha G / \|G\|$. On a alors $\gamma = \alpha / \|G\|$. En d'autre terme, il faudra donc diviser la corrélation par la norme de l'atome auquel elle correspond, pour en déduire le paramètre d'amplitude.

construction du dictionnaire Nous avons donc la possibilité de décomposer le signal analysé sur un ensemble d'atomes arbitraires. Un des points importants est de garder un lien étroit avec les paramètres de synthèse pour générer le dictionnaire. Une des idées sous-tendues par l'élaboration d'une telle méthode était de pouvoir choisir un grain caractéristique dans le signal analysé pour ensuite le transformer suivant les paramètres de synthèse usuels. Dans cette optique, plusieurs choix s'offrent à nous. Le premier est de travailler dans le domaine temporel pour générer les différents grains du dictionnaire. Nous avons notamment expérimenté la transposition du grain caractéristique suivant une échelle de hauteurs spécifiques. Il est bon de remarquer que ce type de transformation va agir autant au niveau de la hauteur du grain que sur sa durée corrélant ainsi fortement les deux paramètres de synthèses. Une des solutions pour contourner cela serait de pondérer le grain par une fenêtre permettant de moduler sa durée mais sommes nous toujours limités à la réduction de la longueur de départ du grain transposé. Une autre approche est de travailler dans le domaine spectral pour la génération des atomes, en effectuant des transformations directement sur le spectre du grain caractéristique comme des translations, des mises à l'échelle,... Mais se pose ici le problème de la resynthèse. En effet, le moteur de synthèse lit dans une table la forme d'onde des grains. Dans notre cas, elle correspond au grain caractéristique et il nous est impossible dans ce schéma de synthétiser des grains issus de transformations spectrales comme l'étirement temporel par exemple. Ainsi, l'élaboration du dictionnaire soulève des problèmes non

triviaux qui étaient difficilement solvables à court terme. Nous n'avons en effet expérimenté que la transposition dans le domaine temporel des grains s'adaptant directement à la synthèse.

3.3.3 Etude de cas

Nous avons effectué des tests de la méthode sur un son de type grattement auquel elle était prévu de s'adapter. En effet, de tels types de sons sont constitués d'une densité moyenne (20 à 30 Hz) de grains complexes. Ces derniers sont le résultat de l'excitation de la structure grattée pouvant ainsi prendre de multiples formes. Ici, il s'agit d'une planche de bois peu dense ayant des modes propres relativement bas. On peut d'ailleurs bien voir sur le spectre du son de grattement représenté en figure 15, la résonance au environ de 215 Hz.

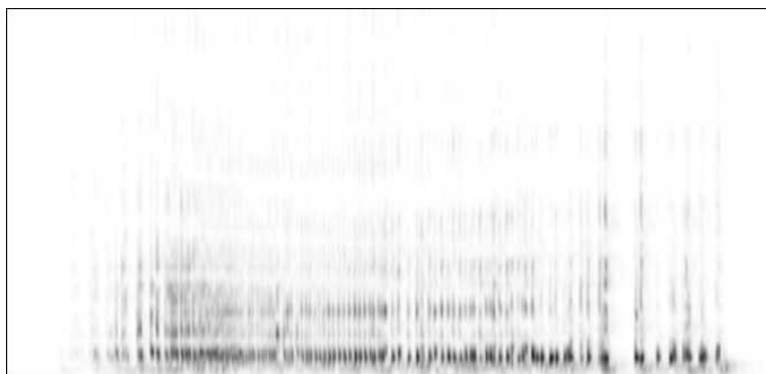


FIG. 15 – Spectrogramme du son de grattement analysé d'une durée de 6 secondes pris entre 0 et 3000 Hz.

Nous avons appliqué l'algorithme avec une TFCT de taille $N = 4096$ et un pas d'avancement $d = 128$ qui donne une résolution temporelle de $r = 128/44100 \simeq 3ms$. Le signal $X[p_{t,f}]$ est construit à partir du module du spectrogramme mais nous avons limité la fréquence maximale permettant de réduire significativement le temps de calcul. En effet, dans notre cas les informations fréquentielles au dessus de 3000 Hz n'étant pas significatives, nous n'avons gardé que les $N_f = \lfloor 3000 * N / F_e \rfloor = 278$ premières fréquences de la TFCT. Le choix de la limite en fréquence doit bien évidemment se faire en accord avec les propriétés du son analysé. Tout d'abord, nous allons effectuer la décomposition sur un dictionnaire comportant un seul grain, ce qui ne peut donner a priori que des résultats peu intéressants mais nous permet de bien cerner le déroulement de l'algorithme. Reste à savoir quel

grain va t'on choisir pour représenter le son d'origine. Ne proposant pas de méthode pour la détermination automatique d'un (ou de plusieurs) grain caractéristique dans le signal analysé, nous avons extrait de manière arbitraire un atome qui perceptivement semblait le plus apte à représenter le son. Les résultats vont ainsi être fortement dépendants de ce choix mais, s'y tenant, on peut tout de même mettre à jour les propriétés intéressantes de la méthode. Le spectrogramme du grain choisi est représenté en figure 16



FIG. 16 – Spectrogramme du grain choisi pour la décomposition, d'une durée de 40 ms, pris entre 0 et 3000 Hz.

On peut alors calculer la corrélation de l'atome avec le signal analysé. Cette fonction est représenté en figure 17.

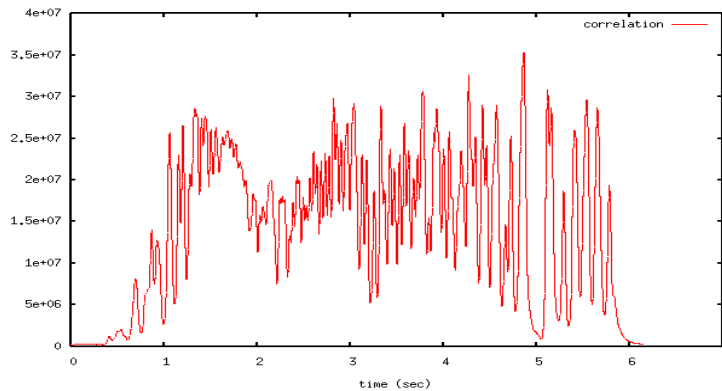


FIG. 17 – Fonction de corrélation entre le grain et le signal de départ.

On pourrait croire que le maximum aux alentours de $t = 5sec.$ correspond à la corrélation du grain du dictionnaire avec son occurrence dans le signal.

Mais il s'agit en fait d'un grain détecté de plus grande énergie et dont la structure est très similaire à celle de l'atome du dictionnaire. Nous avons en effet, extrait l'atome au temps $t = 2.73$ secondes. Rappelons par ailleurs, que la fonction de corrélation ne correspond pas à l'amplitude de l'atome, ceci étant la conséquence de la normalisation des atomes G_k qui permet de comparer la corrélation pour des atomes d'énergie diverse.

Nous avons ensuite effectué des décompositions sur un nombre d'itérations croissant, respectivement de 10, 50 et 100 itérations. Les résultats de ces décompositions sont représentés en figure 18

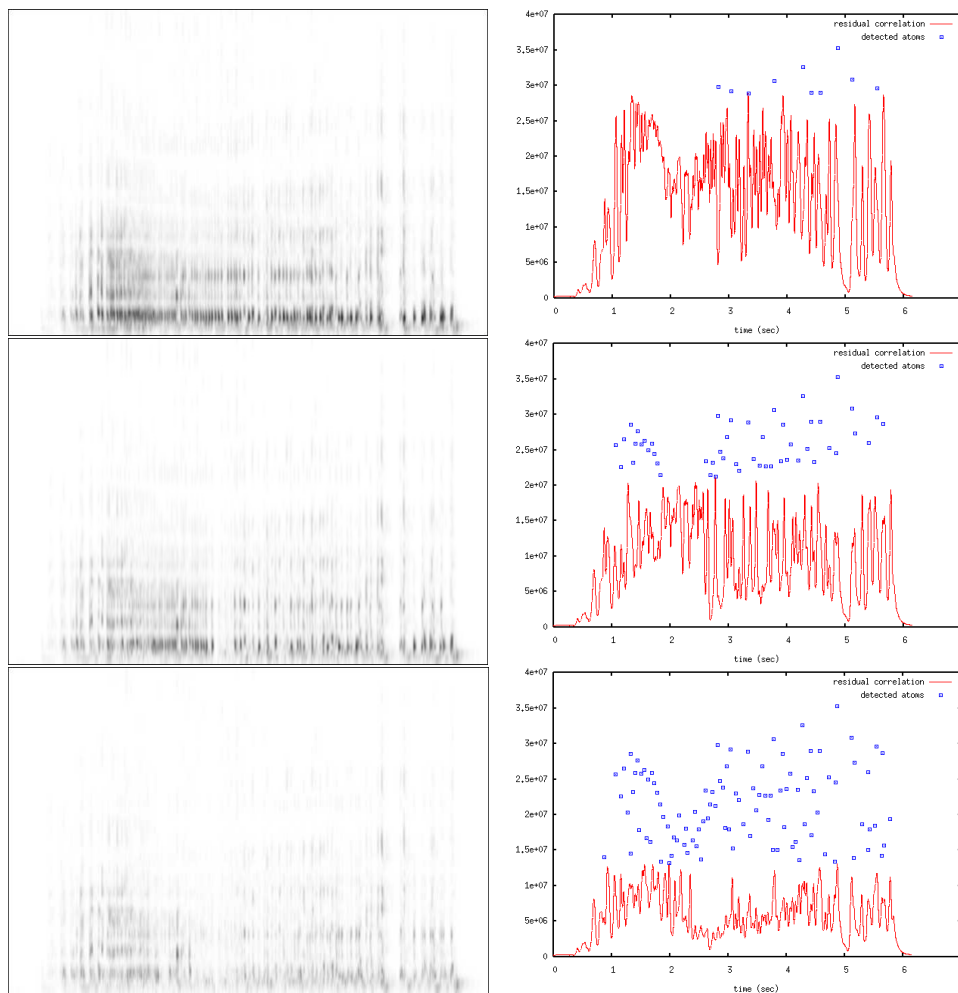


FIG. 18 – Spectres résiduels et fonctions de corrélation à la 10^{ème}, 50^{ème} et 100^{ème} itération de l'algorithme.

Nous avons représenté en figure 18, sur les graphes des fonctions de

corrélation, les maximums à chaque itération correspondant aux atomes détectés. On voit bien sur les résultats de la décomposition l'atténuation de l'énergie du spectre résiduel. L'erreur d'approximation, calculé par $\|R_i\|/\|R_1\|$, passe en effet de 0.91 pour 10 itération, à 0.66 pour 50 et enfin à 0.45 pour 100 itérations. Nous n'avons pas poursuivi la décomposition ayant estimé le nombre de grains dans le signal de départ à une centaine environ. Les atomes détectés par la suite n'aurait fait que tenter de corriger les erreurs aux points de détections antérieurs. En effet, on peut remarquer qu'aux temps où un atome est détecté, la corrélation baisse mais n'atteint jamais la valeur nulle. Cela peut se comprendre par le fait de vouloir représenter notre signal avec un seul et même grain. Mais ceci est d'autant plus étonnant au temps $t = 2.73$ où le grain se retrouve corrélé avec lui même. Ceci s'explique par la résolution temporelle induite dans le calcul du spectrogramme. Les TFCT n'étant calculées que tous les d échantillons (ici $d = 128$), deux mêmes signaux peuvent ne pas avoir le même spectrogramme, pouvant être déphasés par rapport au calcul des transformées. Ces erreurs seront d'autant plus faibles que le pas d est petit, en contre partie d'un temps de calcul multiplié. Le resultat d'amplitude des grains détectés est représenté en figure 19.

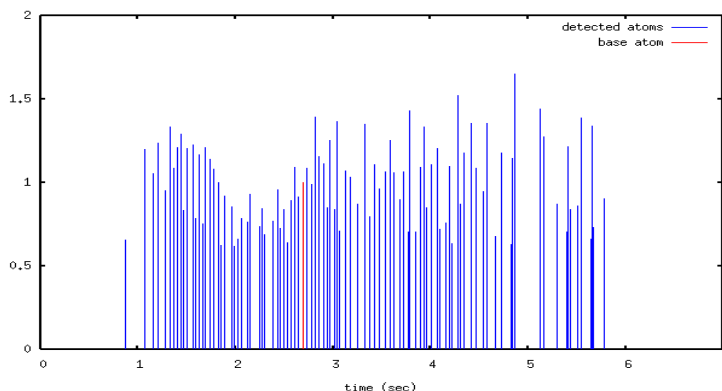


FIG. 19 – Temps et amplitude des grains détectés.

On peut tout de suite voir que le grain ayant servi de base à la décomposition est bien détecté avec une amplitude de 1. En réalité, on trouve une amplitude de 0.99921, erreur négligeable due au problème évoqué ci dessus. Sur les grains forts aux environs de 5-6 secondes, on observe par ailleurs certains doubléments de détection, conséquence d'une mauvaise représentation de ceux ci par le grain choisi. En effet, sur les spectres residuels de la figure 18, on peut voir les composantes aigues persistantes que le grain caractéristique ne peut recouvrir. Il nous est maintenant possible de resynthétiser le signal, le

spectrogramme de cette resynthèse étant représenté en figure 22

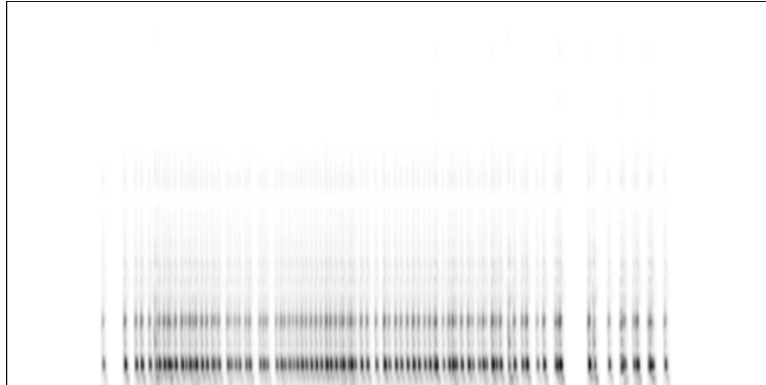


FIG. 20 – Spectrogramme entre 0 et 3000 Hz du son resynthétisé.

Le son résultant est relativement inintéressant par rapport au son original mais il reproduit tout de même fidèlement les propriétés temporelles de ce dernier. On l'associe d'ailleurs plus à un son de grincement. Cela ouvre une parenthèse intéressante, mettant à jour des propriétés de notre catégorisation perceptive de sons bruités. Les sons de grincement sont en effet constitués d'une superposition de grains dont la distribution temporelle peut être comparée à celle des sons de grattements mais à la différence d'une structure spectrale quasi constante pour tous les grains.

Ceci étant, la représentation de notre signal est relativement grossière faisant l'impasse sur la variabilité dans la structure spectrale des grains. Nous introduisons donc les premières expérimentations quand à la construction d'un dictionnaire d'atome. Nous nous sommes basés toujours sur le grain caractéristique avec lequel on a décomposé le son de départ. Il s'agit maintenant de transformer ce grain pour pouvoir, une fois la décomposition effectuée, extraire des paramètres compatibles avec la synthèse. Cela sous-tend l'utilisation du grain de base comme forme d'onde du synthétiseur. Nous pouvons, en accord avec les possibilités de transposition de la synthèse, construire un dictionnaire en modifiant le *pitch* du grain caractéristique dans le domaine temporel, ce qui revient à un changement de fréquence d'échantillonnage.

Au regard de la structure spectrale des grains du signal source, la variabilité de la hauteur n'est pas très étendue. Nous l'avons estimé à un facteur allant d'environ 0.9 à 1.1 par rapport à la hauteur spectrale du grain caractéristique. Nous avons donc construit un dictionnaire comportant 21 grains issue de la transposition, par pas de 0.01, de ce dernier. La décomposition a été faite sur le même nombre de 100 itérations. Les hauteurs

correspondants aux grains détectés sont représentées en figure 21.

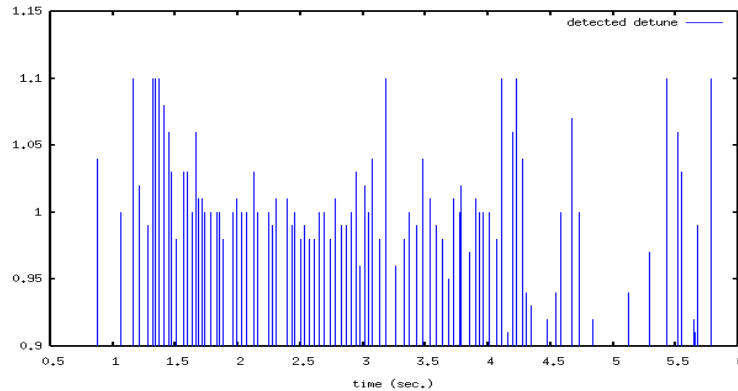


FIG. 21 – Facteur de transposition correspondant aux grains détectés.

On peut vérifier qu’au temps 2,73 secondes la hauteur est bien de 1. correspondant au grain de base. Par ailleurs, on peut voir que la hauteur reste relativement stable autour de la fréquence de référence du grains caractéristique. Cela s’explique par les propriétés du son de départ, qui comme on peut voir sur la figure 15 a une hauteur spectrale assez constante. Par ailleurs, la décomposition sur un tel dictionnaire a-t-il apporté à la qualité de la représentation ? Si l’on en croit la norme du residue, passant de 0.45 pour le dictionnaire ne comportant qu’un atome à 0.44 ici, le gain n’est pas grand et l’utilisation d’une telle technique pour la construction du dictionnaire n’est peut être pas justifiée. Toutefois, à la resynthèse le son est tout de même plus intéressant, perceptivement plus associable à un son de gratement. Son spectre est représenté en figure 22.

Critique Bien que cet exemple soit de l’ordre des premières expérimentations, l’élaboration de cette méthode n’étant pas encore complètement achevée, elle promet tout de même des résultats intéressants pour l’analyse granulaire de sons variés. Il reste toutefois à tester celle ci sur des sons bruités plus complexes et plus denses du point de vue de leur structure. Mais avant cela, de nombreux pans de celle-ci sont encore à expérimenter et à améliorer. La construction du dictionnaire est une des parties les plus sensibles car influençant grandement les résultats de l’analyse. La grande contrainte de cette tâche est d’adapter la structure du dictionnaire à la synthèse granulaire. En effet, pour la transformation du grain caractéristique sur l’espace des paramètres souhaité, on pourrait imaginer des traitements bien plus poussés qu’une simple transposition temporelle. Mais se pose toujours le problème des

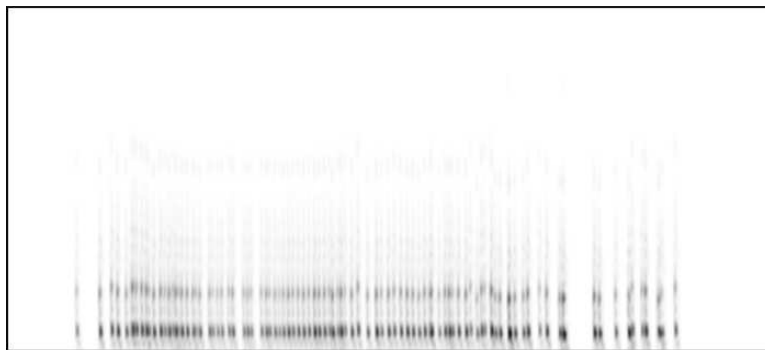


FIG. 22 – Spectrogram entre 0 et 3000 Hz du son resynthétisé suite à une décomposition sur plusieurs grains.

possibilités qu'a la synthèse granulaire pour transformer le grain de base pris en tant que forme d'onde. Peut être faut il étendre la synthèse pour rendre de telles transformations plus flexible? Une des possibilités pour étendre les capacités de représentation du dictionnaire est de prendre plusieurs grains caractéristiques comme bases de la décomposition. Cette approche peut être intéressante même si il ne faut pas perdre de vue l'importance de la compacité de la représentation. Perdre une telle propriété rendrait l'application en synthèse trop dépendante du son resynthétisé rendant par exemple difficiles les traitements type hybridation. Un autre point qu'il faut améliorer sont les critères de choix d'un grain caractéristique et l'évaluation d'une possible routine de détermination automatique. En effet, comme nous l'avons vu, le choix est fait de manière approximative et est loin d'être optimal. Il reste enfin, du point de vue technique, de nombreuses optimisations à effectuer sur le code pour une utilisation ergonomique. Nous avons en effet élaboré l'algorithme sous *octave*, le logiciel scientifique libre type *Matlab*. Cet environnement offre un langage simplifié et une grande précision de calcul, au détriment d'une rapidité d'exécution assez lente. A terme il faudra envisager son implémentation dans des langages permettant un code plus optimisé de type langage C.

4 Synthèse dans MAX/MSP

Ces recherches sur les méthodes d'analyse ont dû s'accompagner de la poursuite du développement des objets MAX/MSP de synthèse granulaire utilisés au GMEM. En effet, la phase de resynthèse impliquait l'ajout de fonctionnalités spécifiques comme l'utilisation d'enveloppes fonctionnelles complexes ou la mise en place d'un générateur de nombre aléatoire suivant des distributions de probabilités arbitraires. Nous allons donc brièvement mentionner les étapes importantes du développement effectué.

4.1 Générateur de nombre aléatoire de loi arbitraire

Dans MAX/MSP, les objets permettant la génération aléatoire ne peuvent généralement fournir que des variables équiprobables. Or, il était intéressant pour le contrôle des paramètres de la synthèse granulaire de pouvoir spécifier des distributions plus évoluées. L'idée est alors de définir des fonctions de distribution sous la forme de liste de valeurs, chaque valeur correspondant à la probabilité du tirage de l'index qui lui est associé. Cela peut se formaliser ainsi. Soit la liste $\{p_k\}$ de longueur K correspondant à la fonction de probabilité telle que :

$$\sum_{k=0}^K p_k = 1$$

On veut :

$$P(Y = k/K) = p_k$$

Avec Y variable aléatoire de sortie. Soit X variable aléatoire équiprobable tel que $X \in [0; 1.]$. Or si $(a, b) \in [0; 1.]^2$ et $a < b$:

$$P(a < X < b) = b - a$$

Introduisons par ailleurs

$$S_k = \sum_{n=0}^k p_n$$

On a donc :

$$P(S_k < X < S_{k+1}) = p_{k+1}$$

et ainsi peut on en déduire le calcul de Y :

$$Y = (\arg_{k \in [0; K]} [S_k < X < S_{k+1}]) / K$$

En d'autres termes, il faut créer la liste des sommes successives de la liste de probabilité et ensuite comparer une variable aléatoire équiprobable avec

chacun des éléments. L'index du premier élément plus grand que la variable correspond à la sortie recherchée. Le patch MAX réalisant un tel traitement est représenté en figure 23. Techniquement, l'objet *vexpr* convertie la liste S_k

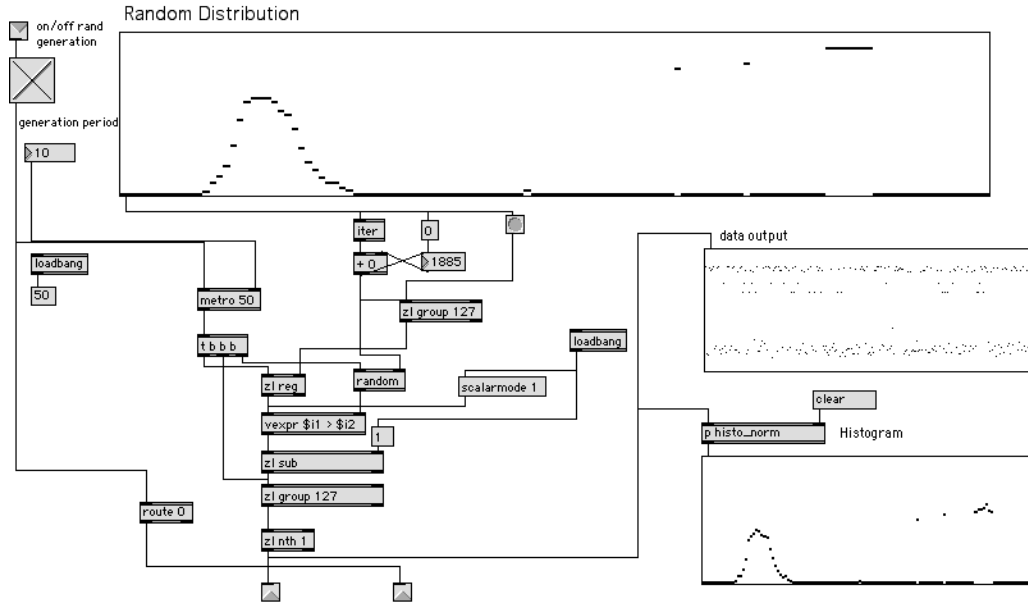


FIG. 23 – Patch MAX pour la génération de nombres aléatoires de loi arbitraire sous forme messages.

en liste de valeur binaire selon le résultat de la comparaison. *zl sub* renvoie en sortie les index des occurrences de valeurs vrais (ou = 1) et *zl nth* prend le premier élément de cette nouvelle liste. On peut spécifier la fréquence de génération des valeurs aléatoires par la modification de la période de l'objet *metro*. On peut descendre jusqu'à la limite dictée par l'horloge message de MAX qui est d'environ 1 ms.

Cet objet offre des résultats très satisfaisants pour des fréquences de générations moyenne ($f < 100$ Hz). Mais franchi ce seuil, on obtient des irrégularités non négligeable dans la fréquence réelle de génération. Ceci est dû à la gestion des messages dans MAX, ce dernier n'assurant pas une grande exactitude temporelle dans la propagation des messages. Or on veut pouvoir, avec le module de synthèse créer des flux granulaires dense (de densité > 1000 grains/secondes) avec la possibilité de spécifier pour chacun des grains des paramètres stochastiques propres. L'utilisation de cette version du générateur introduit dans ce cas des artefacts audibles, la fréquence du générateur ne suivant pas celle de déclenchement des grains. La seule solution à ce problème

est d'utiliser un générateur de type signal pouvant théoriquement monter jusqu'à la fréquence d'échantillonnage. Nous avons donc développé un objet externe en langage C pour réaliser cette tâche. Le patch d'aide de cet objet est représenté en figure 24.

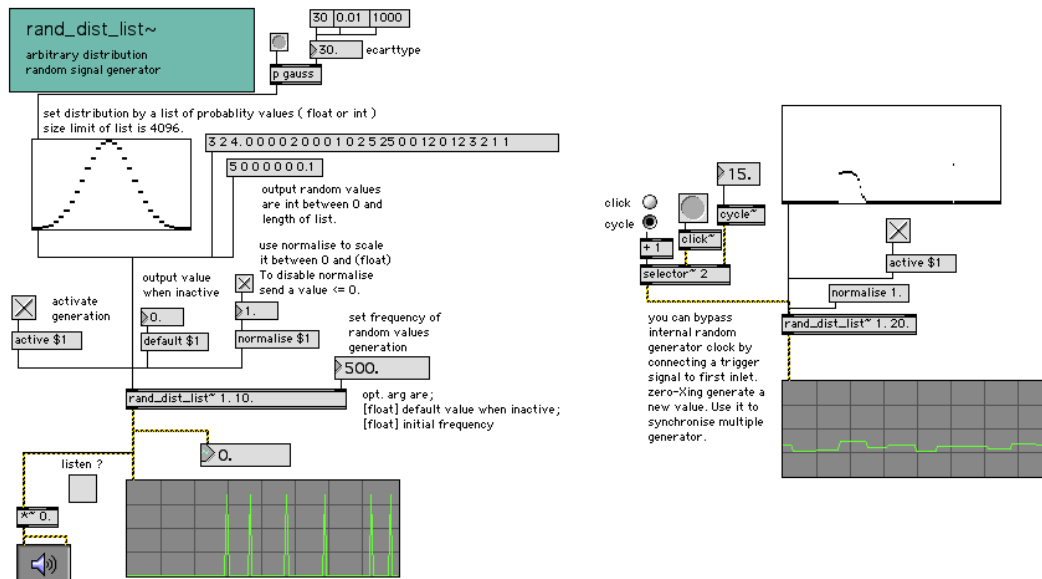


FIG. 24 – Objet externe pour la génération de nombres aléatoires de loi arbitraire sous forme signal.

Cet objet suit le même schéma que la version message prenant une liste de probabilité en entrée. Il a une horloge interne mais peut aussi être contrôlé par un signal externe. En connectant un signal sur la première entrée, il va générer une nouvelle valeur à chaque croisement par zero de ce dernier. Ceci peut être très utile pour la synchronisation de plusieurs générateurs. On peut ainsi avec ce système avoir un contrôle précis des paramètres fournis au module de synthèse.

4.2 Objet de synthèse granulaire avec enveloppe sous forme de table

L'objectif du GMEM est de créer 9 objets de synthèse granulaire différents selon leurs applications spécifiques. Les objets de synthèse alors implémentés et opérationnels sont ceux dont la forme d'onde est basée sur une table *buffer~* (colonne centrale de la figure 25). Il ont été implémentés principalement par Laurent Pottier et Loic Kessous. Les objets *lnGranul~* et


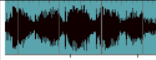


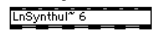

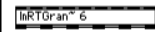

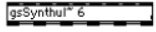
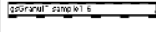
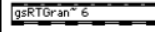

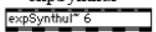
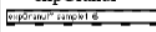
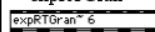
envelopes \ waveforms	 sine waveform	 sample	 audio input
 linear	lnSynthul~ 	lnGranul~ 	lnRTGran~ 
 gaussian	gsSynthul~ 	gsGranul~ 	gsRTGran~ 
 cosine	expSynthul~ 	expGranul~ 	expRTGran~ 

FIG. 25 – Objets MAX/MSP de synthèse granulaire

cosGranul~ travaillent avec des enveloppes à segments linéaires ou demi-cosinus. *gsGranul~* quand à lui propose une enveloppe gaussienne générée en temps-réel. Ils étaient tout trois déclinés en deux versions : une acceptant des paramètres sous forme messages et l’autre sous forme signal. Il pouvait être intéressant de définir des enveloppes arbitraires sous forme de *buffer~* externe. Cela pouvait notamment remplacer les fonctions d’un *gsGranul~* coûteux en temps de calcul. Nous nous sommes basés sur le code C des objets existants pour élaborer un tel objet.

Le seul paramètre influenceant l’enveloppe est alors le paramètre de longueur *length*. Nous avons veillé à effectuer une interpolation linéaire pour la lecture de la table, ceci évitant les artefacts audibles d’*aliasing*. Par ailleurs, les grains sont déclenchés par le passage par 0 d’un signal *trigger*. Il est ainsi impossible de générer plusieurs grains exactement synchrones. Nous avons donc ajouté une fonctionnalité appelée *gpacket mode* permettant de définir des groupes de grains, déclenchables par le signal *trigger*.

Pour les mêmes raisons évoquées plus haut, même si il est possible avec un tel objet de déclencher un grand nombre de grains par secondes, les variations des paramètres ne peut suivre des cadences plus grande qu’environ 1000 messages/seconde. Nous sommes donc limités par l’horloge message de MAX. En réalité, rien ne nous empêche de créer des structures denses avec génération stochastiques des paramètres mais le signal résultant est alors fortement corrélée avec la fréquence de mise à jour de ceux ci. D’où l’intérêt, de l’extension de cet objet à une version acceptant des paramètres de type signal. Nous avons donc développé un tel objet appelé *bufGranul~*.

Du point de vue de son fonctionnement, les signaux paramètres sont échantillonnés au temps exact (“temps signal” de résolution sa fréquence

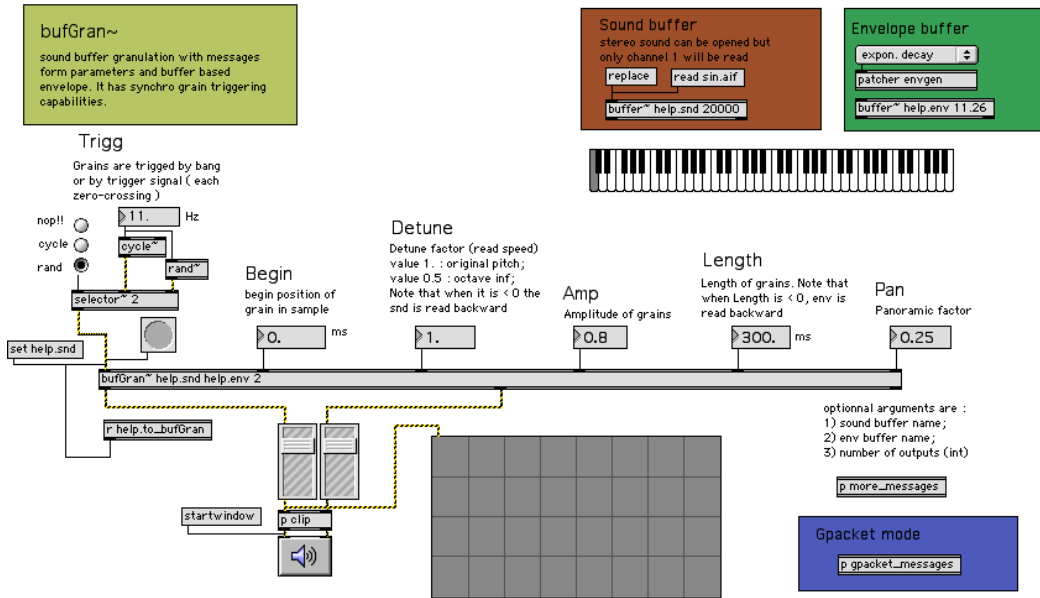


FIG. 26 – Patch d’aide de l’objet *bufGran~*

d’échantillonnage) où le déclenchement d’un grain a lieu, déclenchement qui est d’ailleurs contrôlé par signal. Ceci nous donne ainsi la possibilité d’une part d’une fréquence de mise à jour de paramètres jusqu’à la fréquence d’échantillonnage. D’autre part cela permet d’avoir un contrôle précis sur les paramètres d’un grain particulier “noyé” dans un flux de n’importe quelle densité. Ceci est notamment intéressant pour faciliter à terme l’implémentation des dépendances entre les paramètres, pouvant alors lier de manière fiable l’évolution de plusieurs paramètres. L’association entre l’objet *ran_dist_list~* de génération de signal aléatoire et *bufGranul~* donnent des résultats bien plus intéressants pour la synthèse de flux granulaires denses, n’ayant plus d’artefacts liés à l’horloge message. Ils constituent ainsi un bon environnement de tests pour la resynthèse de sons bruités issus d’analyse.

4.3 Vers une interface graphique de contrôle...

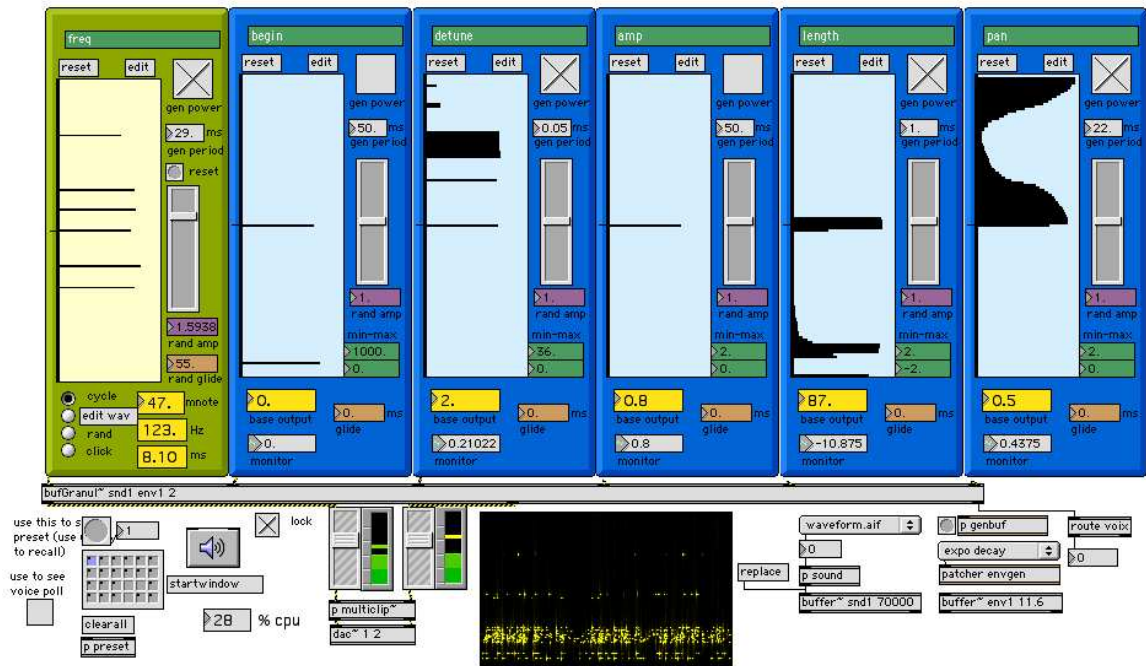


FIG. 27 – Interface de contrôle de l'objet *bufGranul~* de synthèse granulaire.

5 Conclusion et perspectives

5.1 Bilan

L'amorce de ce projet de recherche sur l'application de la synthèse granulaire au sons bruités donne des résultats encourageants autant au niveau de l'analyse comme de la synthèse. En effet, l'amélioration des objets de synthèse en regard de cette application spécifique sont un facteur motivant pour approfondir les recherches en analyse. Empiriquement, on arrive à recréer des sons bruités de types très variés qui sont perceptiblement probants. L'ajout des fonctionnalités de generations aléatoires de loi arbitraire y est pour beaucoup dans ces résultats. Le monde physique est en effet régi par des phénomènes complexes, dont la modélisation par des fonctions probabilistes simples ne peut être que réductrice. Avec des fonctions aléatoires complexes, on arrive à des résultats qui sonnent perceptiblement beaucoup plus naturels. Les classes de sons ainsi créés sont très diverses comme des bruits de gouttes d'eau, de grattements, de tremblements, ou des sons plus de l'ordre du bruits quasi-constant comme ceux de mers, de vents,... Ceci nous montre les possibilités très prometteuses qu'a la synthèse granulaire dans cette application.

Concernant l'analyse, les différentes méthodes évaluées ont des intérêts divers et complémentaires. Pour des sons bruités dont les grains sont séparables temporellement et ont des variations simples, la méthode par segmentation peut être une technique facile à mettre en oeuvre et pouvant donner des résultats satisfaisants. La méthode Matching Pursuit quand à elle est très efficace pour la décomposition en grains simples et par une possible configuration plus poussée peut s'adapter à des cas divers. Spectral Matching Pursuit, bien que son élaboration ne soit pas encore aboutie, promet des résultats exploitables sur de larges classes de sons bruités permettant la représentation de ceux-ci par l'évolution paramétrique de grains complexes.

Ces constats permettent d'entrevoir à terme la possibilité d'une catégorisation de sons bruités par leurs paramètres applicables en synthèse granulaire.

5.2 Optimisations et Perspectives

Outres les optimisations envisagées pour chaque méthode précédemment évoquées, un des points majeurs pour l'évolution du modèle est la gestion de la corrélation des paramètres de synthèse entre eux. Il semble en effet que cela soit une donnée importante pour la modélisation de sons bruités. les paramètres extraits résultent de phénomènes physiques complexes dont les propriétés influencent ces premiers de manière globale, les rendant ainsi interdépendants. D'un point de vue technique, ceci implique d'une part un trai-

tement spécifique des données analysées déterminant la nécessité de corrélérer tel ou tel paramètre et avec quelles modalités. D'autre part, il faut alors mettre en place, au niveau de la synthèse, un système de génération de paramètres contraint par les règles de corrélation ainsi déterminées. On notera que cette dernière étape sera facilitée par la gestion des paramètres de type signal et la fonction de synchronisation implémentée dans les objets *rand_dist_list*.

Du point de vue de la synthèse, il faudra aussi à terme pouvoir gérer l'évolution des lois stochastiques régissant les paramètres et cela de manière ergonomique. Une solution simple dans MAX pourrait être l'utilisation d'un séquenceur de messages mettant à jour les courbes à des temps définis. Mais on voudrait ici un outils plus malléable pouvant effectuer des interpolations entre différentes fonctions de probabilités. Cela permettrait d'une part d'avoir une évolution des paramètres de mise en oeuvre plus aisée pour le compositeur et d'autre part d'explorer l'espace sonore entre plusieurs jeux de paramètres issus d'analyses (hybridation).

5.3 Apports Personnels

Ce travail a été pour moi très enrichissant à de multiples niveaux. Le GMEM a en effet été une structure idéale pour la réalisation de mon stage ATIAM, étant un lieu d'échange permanent entre recherche scientifique et création artistique. Cela m'a permis d'appliquer les notions enseignées lors de la formation ATIAM, en ayant toujours à l'esprit l'utilisation finale des outils par des compositeurs. Mes travaux ont ainsi oscillé entre recherche théorique et application pratique, deux domaines qui sont pour moi complémentaires. De plus, ce sujet de recherche, au début de mon stage, était relativement nouveau et encore peu exploré. Cela a rendu le travail réellement captivant, devant imaginer et étudier des méthodes à partir de peu d'existant, une grande marge de manoeuvre m'ayant été accordée. Pour toutes ces raisons, je remercie Laurent Pottier et Raphaël De Vivo pour m'avoir permis de participer à un tel projet.

Références

- [1] R. Gribonval et al. Analysis of sound signals with high resolution matching pursuit. *Proceedings of TimeFrequency Time-Scale Symposium*, 1996.
- [2] Xavier Rodet et Gerald Bennett. Synthèse de la voix chantée par ordinateur. *Conférences des Journées d'Etude, Festival International du Son, Paris*, 1980.
- [3] Michael M Goodwin and Martin Vetterli. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Trans. on Signal Proc*, 1998.
- [4] R. Gribonval. Fast ridge pursuit with a multiscale dictionary of gaussian chirps, 2002.
- [5] R. Gribonval, X. Depalle, E. Bacry, and S. Mallat. Sound signal decomposition using a high resolution matching pursuit. *Proceedings of International Computer Music Conference (ICMC'96). Clear Water Bay, Hong-Kong*, 1996.
- [6] P. Hanna, A. Beurive, and M. Desainte-Catherine. Real-time noise synthesis with control of the spectral density. *Proceedings of the Digital Audio Effects Workshop (DAFX'02, Hamburg, Germany)*, 2002.
- [7] P. Hanna and M. Desainte-Catherine. Statistical approach for sound modeling. *Proceedings of the Digital Audio Effects Workshop (DAFX'00, Verona, Italy)*, 2000.
- [8] S. Jaggi, W. Carl, S. Mallat, and A. Willsky. High resolution pursuit for feature extraction. *Technical report, MIT*, 1995.
- [9] Damian Keller. Touch'n'go : Ecological models in composition. *url : www2.sfu.ca/sonic-studio/srs/EcoModelsComposition/Title.html*, 1999.
- [10] Rocha M. Les techniques granulaires dans la synthèse sonore. *thèse de doctorat en Musique, Université Paris VIII*, 1999.
- [11] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, 1993.
- [12] Timothy Opie. An overview of granular synthesis and the techniques involved. *<http://www.granularsynthesis.live.com.au/hthesis/>*, 1994.
- [13] Laurent Pottier. Gmu - an integrated microsound synthesis system. *Proceedings of the Computer Music Modeling and Retrieval Conference, Montpellier (France)*, 2003.

- [14] Curtis Roads. Asynchronous granular synthesis. *Representations of Musical Signals*, MIT Press, 1991.
- [15] Curtis Roads. *Microsound*. MIT Press, 2001.
- [16] Eric D. Scheirer. Tempo and beat analysis of acoustic musical signals. *Acoustical Society of America*, 1997.
- [17] Barry Truax. Discovering inner complexity : Timeshifting and transposition with a real-time granulation technique. *Computer Music Journal* 18, 1994.
- [18] Damian Keller Barry Truax. Ecologically-based granular synthesis. *ICMC*, <http://www.sfu.ca/dkeller/EcoGranSynth/EGSpaper.html>, 1998.