

DU GRAIN À LA MUSIQUE : MANIPULATIONS EN TEMPS RÉEL

Isabel Pires

IRCAM – 1, pl. Igor Stravinsky,
75004 Paris, France
CICM - Université Paris VIII
MSH Paris Nord
Isabel.Pires@ircam.fr ; isabel_pires@excite.com

Norbert Schnell

IRCAM – 1, pl. Igor Stravinsky,
75004 Paris, France
Norbert.Schnell@ircam.fr

RÉSUMÉ

Dans cet article, nous présenterons une étude comparative de quelques techniques de manipulation du signal comme la Transformée de Fourier à Court Terme, la Synthèse Granulaire Asynchrone et PSOLA, notamment dans le cadre de leur utilisation en temps réel. Nous présenterons succinctement le fonctionnement de ces trois méthodes dans l'environnement Max / MSP avec l'utilisation de la librairie GABOR.

La manipulation des paramètres propres à chaque méthode a des conséquences sur les variables musicales, et l'utilisation de ces méthodes implique des corrélations entre les paramètres techniques du traitement du signal et des attributs musicaux spécifiques. Ainsi, nous essayerons d'explorer les espaces des paramètres et de la perception musicales entre le temps, la période et le spectre (les rythmes, les fréquences et les timbres) en comparant les différentes méthodes en vue de leur croisement dans un système de traitement commun.

Mots-clés : Synthèse Granulaire Asynchrone, PSOLA, Transformés de Fourier à Court Terme, Convolution, Temps Réel.

1. INTRODUCTION

Les méthodes de granulation du signal consistent à produire des textures sonores en manipulant des grains de son (échantillons plus ou moins courts), par transposition, dilatation et compression temporelle, entre autres choses. Le grain constitue une entité sonore « atomique ». Dans ce sens, beaucoup des écrits sur la synthèse granulaire font référence à Denis Gabor, qui évoque en premier l'idée d'un « quantum acoustique ». Même si la synthèse granulaire ne peut pas revendiquer une fidélité aux idées de Gabor, c'est dans les écrits de Gabor qu'on trouve pour la première fois une description à la fois mathématique et acoustique d'une « particule » sonore. Ses descriptions nous aident à comprendre et à anticiper l'effet des traitements des entités sonores très brèves, se trouvant aux bords de la perception des périodes, pour la synthèse granulaire mais aussi pour d'autres techniques de traitement comme PSOLA ou *Chant*. Nous voudrions aussi inclure dans ces réflexions des techniques basées sur la transformée de Fourier à court terme ainsi que d'autres techniques qui s'appuient sur la superposition de trames, grains ou formes d'ondes.

Nous aborderons ces techniques à travers l'utilisation de GABOR, une librairie basée sur une notion généralisée de la synthèse granulaire. Cette librairie introduit un paradigme de traitement de signal basé sur l'idée d'un flux de particules sonores librement manipulables au niveau des durées et des vitesses de traitement.

La librairie Gabor permet le traitement de n'importe quel type de particules sonores (des grains, des formes d'onde élémentaires, des fenêtres temporelles) en ajustant sa vitesse d'extraction et sa position temporelle à la fréquence, la texture, le rythme, ou encore une vitesse spéciale de traitement exigée par un algorithme particulier.

Nous décrivons d'abord les techniques de Transformée de Fourier à Court Terme (STFT – Short Time Fourier Transform) [6], Synthèse Granulaire Asynchrone (AGS – Asynchronous Granular Synthesis) [6], et PSOLA (Pitch Synchronous OverLap-Add) [4], ceci dans le cadre d'une exploitation en temps réel. Ensuite, nous décrivons certaines manipulations, de paramètres temporels, fréquentiels et spectraux, relatives respectivement, à la STFT, à l'AGS et à PSOLA par rapport aux variables musicales du timbre, du rythme et de la hauteur.

2. GABOR ET LES « PARTICULES SONORES »

La librairie Gabor offre diverses méthodes pour couper un flux sonore continu en un flux de vecteurs en utilisant divers schèmes temporels et implémente divers algorithmes d'analyse et de synthèse pour le traitement de ces vecteurs. Chaque vecteur de ce flux est la description d'une particule sonore. La reconversion des flux de vecteurs en un nouveau flux sonore continu se fait à l'aide d'une méthode générale de superposition-addition.

En ce moment, Gabor offre deux modules différents de ligne à retard permettant de faire le traitement des particules du flux sonore tout en assurant la vitesse et la fluidité nécessaires au temps réel. Le module `gbr.dline~` fonctionne comme une ligne à retard qui cumule les derniers échantillons du flux sonore où l'on va extraire des particules sonores sous la forme de vecteur, tandis que le module `gbr.drain~` emmagasine les échantillons traités à faire suivre vers la sortie. Ces modules fonctionnent comme un des interfaces entre le flux sonore et le flux de vecteurs et vice-versa.

Gabor offre des modules spécialement conçus pour le traitement des flux sonores de façon spécifique, spécialement dirigés à des techniques particulières de traitement du signal : `gbr.copy` et `gbr.dline~` pour la synthèse granulaire ; `gbr.slice~` pour les méthodes STFT ; `gbr.psy~` pour des méthodes du type PSOLA. Nous retournerons à ces modules plus loin.

3. TECHNIQUES GRANULAIRES STFT, AGS et PSOLA

3.1. Transformé de Fourier à Court Terme

La *Transformé de Fourier à Court Terme* (STFT) est une méthode qui permet le traitement du signal en le décomposant en représentations spectrales successives à travers un fenêtrage temporellement défini, mais très court. De ce fait la dénomination à *court terme*. La fonction de fenêtrage est une enveloppe temporelle d'une durée généralement comprise entre 1ms et 100ms.

Chaque segment sonore analysé produit un ensemble de données – appelé image – contenant une description des amplitudes ainsi que des valeurs de phase des composants fréquentiels du spectre. L'analyse spectrale de chaque segment sonore permet l'obtention d'une représentation descriptive de l'évolution du contenu spectral du son au cours du temps.

L'analyse / re-synthèse du signal, en utilisant des méthodes du type STFT, est habituellement réalisée à l'aide de techniques de superposition-addition de fenêtres / images. L'utilisation de techniques de superposition-addition dans les méthodes du type STFT est due à la nécessité de capter et de reconstituer le signal sans perdre trop d'information.

Au niveau de l'analyse, les fenêtres utilisées pour couper le signal sont généralement du type *gaussien* car cette forme réduit la distorsion qui peut apparaître au moment du fenêtrage. Ainsi, la superposition de fenêtres d'analyse permet de réduire, ou même d'éliminer, la perte d'information relative au signal qui pouvait advenir à cause de la courbe de la fonction de fenêtrage. Le même se passe au moment de la re-synthèse afin reconstituer le signal dans son intégralité. Toutefois, cette méthode de superposition-addition, si utilisée de façon stricte, peut conduire à des limitations du point de vue des transformations musicales. Ainsi, le contrôle des paramètres de cette méthode est très important pour la musicalisation de la manipulation sonore.

La STFT est à la base de différentes techniques de traitements temps-fréquence-spectre comme le filtrage par convolution, la synthèse croisée, la transposition ou les modifications de l'échelle temporelle. De point de vue de la synthèse granulaire, les techniques basées sur la STFT imposent souvent une contrainte particulière de cohérence des phases des composants du signal (des grains successifs) ; c'est cette cohérence qui assure une continuité perceptive du signal résultant. C'est surtout dans les fondements théoriques du Vocodeur de Phase (PV- Phase Vocoder) [6] [7] que ces contraintes sont formulées de manière stricte. Ces contraintes s'opposent en quelque sorte aux effets granulaires d'une technique

de type « overlap-add ». Malgré cela, le paramétrage du vocodeur de phase nous renvoie aux préoccupations impliquées dans la granulation du son, c'est-à-dire aux paramètres de traitement des entités sonores « atomiques », aussi bien en termes de résolution temporelle que fréquentielle.

Dans le sens de optimiser une interface pour les méthodes du type STFT de façon à minimiser les contraintes énoncés, la librairie Gabor compte le module `gbr.slice~` qui permet la fenêtrage des flux sonores en vecteurs de taille spécifique et parfaitement synchrones au niveau de la phase. L'échantillonnage, en utilisant `gbr.slice~`, engendre un flux sonore en vecteurs superposés, la superposition, définit en échantillons, remplit toutes les conditions spécifiques des techniques d'analyse / re-synthèse du type STFT.

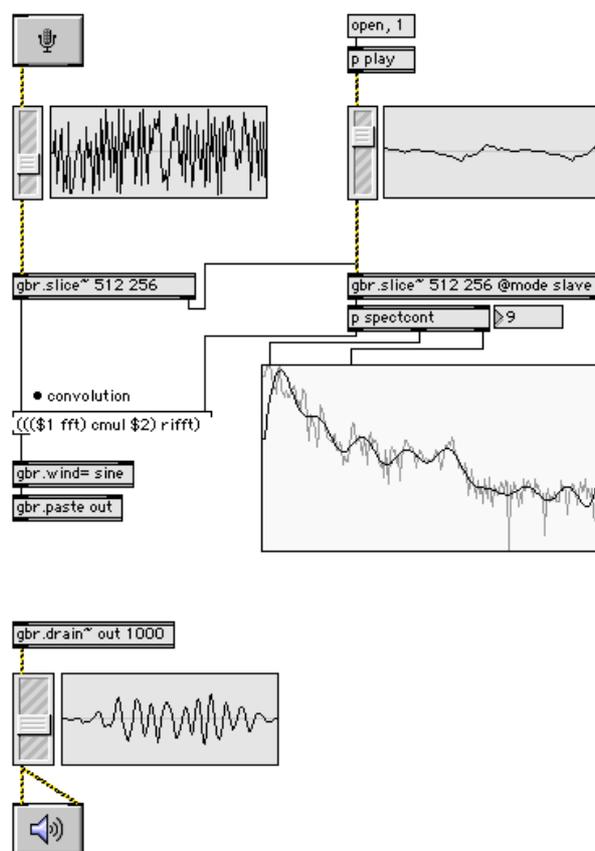


Figure 1. Exemple de convolution basé sur le méthode STFT en utilisant le module `gbr.slice~`.

La figure 1. montre un exemple l'application de la convolution en utilisant des modules de la librairie Gabor. La manipulation des paramètres des diverses techniques de traitement sonore basées sur la STFT permet d'obtenir des transformations au niveau du timbre et de la texture. L'ajustement de ses réglages dépend à la fois des caractéristiques du son traité et du type de résultat recherché.

3.2. Synthèse Granulaire Asynchrone

La Synthèse Granulaire Asynchrone (AGS) permet une grande liberté dans la manipulation des paramètres du grain dans la production d'une texture sonore. Ce degré de liberté est traditionnellement conditionné par les contraintes du temps réel à cause de son besoin d'un « chunk » de mémoire où est stocké le fichier de son sur lequel est appliquée la granulation. Cependant, l'utilisation de GABOR vient réduire cette contrainte permettant l'ouverture d'une fenêtre qui « glisse » sur le signal qui arrive en temps réel. La taille de cet espace est réglable, permettant ainsi la manipulation granulaire du son.

Les paramètres de la AGS contrôlables en temps réel sont :

- ❑ Taille de la fenêtre – durée du grain (de 1 à 100 ms habituellement)
- ❑ Type de fenêtre – forme d'onde de l'enveloppe de fenêtre
- ❑ Distance temporelle de fenêtrage
- ❑ Densité :
 - Dispersion spatiale
 - Dispersion temporelle
 - Dispersion fréquentielle
- ❑ Enveloppe dynamique de la masse sonore
- ❑ Variation fréquentielle de la masse sonore

À l'exception de ceux qui se rapportent à la masse sonore, nous remarquons que les paramètres énoncés ont un caractère principalement temporel, lié à la durée – soit par taille du grain soit par la densité (localisation / superposition dans l'axe temporel). Il est précisément ce caractère de temporalité qui peut conditionner une certaine liberté dans les manipulations en temps réel.

Pour la synthèse granulaire (asynchrone ou pas), Gabor permet de couper des grains dans le flux sonore en utilisant d'abord le module `gbr.dline~` comme réservoir des derniers échantillons en coordination avec le module `gbr.copy` dans lequel est indiquée la durée de chaque grain. Ainsi les modules `gbr.dline~` et `gbr.copy`, fonctionnant ensemble engendrent les grains sonores, tandis que le fonctionnement des modules `gbr.paste` et `gbr.drain~` engendre la re-synthèse du signal sonore par superposition-addition.

L'utilisation du module `gbr.copy` permet, à n'importe quel instant temporel, d'effectuer la copie du vecteur descriptif d'une particule sonore dans la ligne à retard (`gbr.dline~`), de lui traiter et de lui reconstituer en particule sonore d'un flux continu à l'aide du module `gbr.paste`.

Il faut encore déterminer la localisation de l'extraction du grain dans le flux sonore qui arrive, ainsi que son repositionnement dans le flux sonore re-synthétisé. Tantôt la position spécifique de chaque grain, tantôt sa taille, et la densité granulaire du flux sonore reconstitué, sont des variables manipulables en temps réel.

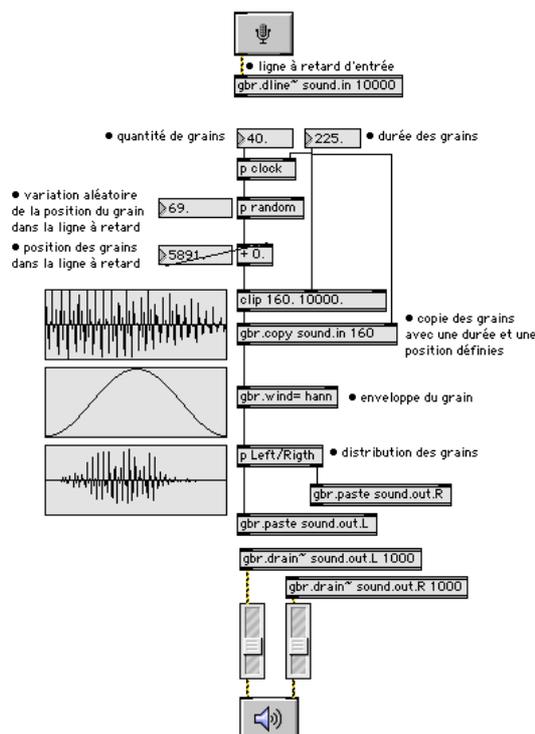


Figure 2. Exemple de synthèse granulaire asynchrone en utilisant les modules `gbr.dline~` et `gbr.copy`.

La figure 2. montre un exemple très simple de l'application de synthèse granulaire asynchrone utilisant des modules de la librairie Gabor.

Les vecteurs extraits au flux à l'aide du module `gbr.copy` sont temporellement autonomes par rapport au moment de son engendrement. C'est-à-dire, un vecteur, copié dans la ligne à retard, peut être utilisé comme particule autonome pour générer un flux de vecteurs d'une longueur temporelle et de caractéristiques différentes de celles du vecteur copié à l'origine. Ainsi, à l'aide d'un seul grain est possible d'engendrer un flux sonore, une texture granulaire.

3.3. PSOLA

PSOLA est une méthode « d'analyse et synthèse par superposition-addition de fenêtres synchrones à la période fondamentale du signal » [4]. Cette méthode, fonctionnant sur des signaux monophoniques, a été spécialement conçue pour la synthèse de la voix, et permet la manipulation du son tout en préservant ses caractéristiques structurelles, notamment les caractéristiques spectrales (formants, transitoires d'attaque). Le fenêtrage synchrone, tant à l'analyse qu'à la synthèse, permet de contrôler aussi bien le déroulement de l'axe temporel que la hauteur du signal, c'est-à-dire, les aspects fréquentiels et temporels.

Cette méthode de fenêtrage, centrée sur les périodes

fondamentales du signal, décompose le signal dans des formes d'onde élémentaires et permet sa manipulation musicale sans changer nécessairement sa structure interne. C'est-à-dire, sans changer ni ses structures spectrales ni ses structures temporelles.

PSOLA peut être perçue comme une méthode de synthèse granulaire dans le sens que chaque grain correspond à la forme d'onde élémentaire instantanée qui a été extraite du signal d'entrée.

L'analyse doit séparer les parties périodiques (harmoniques) des parties non-périodiques du signal (transitoires d'attaque, zones bruiteuses)[4]. Souvent la synthèse PSOLA est complétée avec une synthèse granulaire asynchrone, pour les parties bruitées du son.

Les caractéristiques propres à cette méthode d'analyse / synthèse, basée principalement sur la manipulation des parties périodiques du signal, donnent de l'aisance dans le traitement des paramètres fréquentiels du son.

Concernant l'analyse / re-synthèse PSOLA, avec ses caractéristiques et contraintes spécifiques, Gabor inclus le module `gbr.psy~`. Ce module coupe le flux sonore d'entrée en vecteurs des formes d'onde élémentaires successives des parties périodiques du son et engendre des grains de période et durée asymétrique (en dépendant du contenu spectral du son analysé) des parties non-périodiques du son.

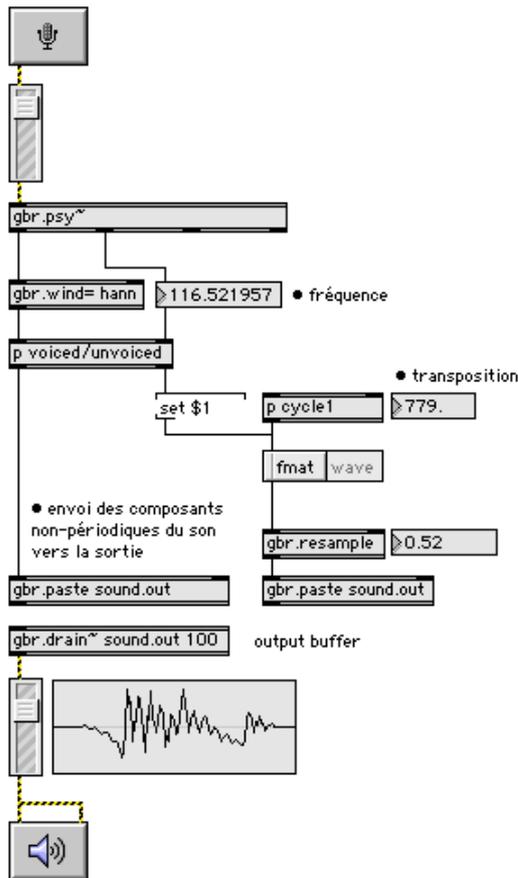


Figure 3. Exemple d'analyse / re-synthèse PSOLA en utilisant le module `gbr.psy~`

4. MANIPULATION DES VARIABLES MUSICALES

La manipulation en temps réel des aspects énoncés pour chacune des trois méthodes citées, trouve des correspondants dans la composition musicale aussi bien que dans la perception auditive. La composition musicale a tendance, par principe, à détourner les paramètres « techniquement conseillés » en faisant basculer les résultats attendus, ou considérés comme caractéristiques d'une certaine technique, en des effets sonores musicalement signifiants, même s'ils sont inattendus.

Dans ce contexte, les axes de corrélation entre des indices techniques et musicaux, présentés dans ensuite, démontrent qu'ils ne sont pas des champs de travail fermés sur eux-mêmes. La malléabilité des techniques décrites ici, par rapport aux résultats sonores / musicaux souhaités, ainsi que la fluidité entre les divers effets sonores passibles d'être obtenus avec chacune d'entre elles, démontre la grande capacité d'intégration de ces techniques dans champ musical, et particulièrement par rapport au temps réel.

4.1. Temporalité : durée – rythme

La manipulation en temps réel des aspects temporels, dans les techniques de traitement granulaire du son comme la AGS, conduit, en premier plan, au traitement musical de la variable rythme. Ces manipulations temporelles peuvent aller du simple engendrement d'impulsions à la production de trames sonores aux textures complexes. Néanmoins, d'autres aspects comme la hauteur et le timbre peuvent être affectés par le traitement des paramètres du signal liés à la temporalité.

En fait, le traitement des paramètres temporels du signal a des conséquences musicales et sonores au niveau des notions de rythme, hauteur et timbre, mais, les sensations perceptives de fusion et scission, discrètes aussi bien que continues, y sont également concernées. C'est dans ce contexte que la perception auditive devient déterminante dans la caractérisation des aspects musicaux par rapport aux variables techniques.

Or, les sensations de fusion et scission, discrètes et continues, sont directement liées à la caractérisation perceptive des aspects (indices) musicaux. Par exemple, si, en utilisant la méthode AGS, les paramètres choisis produisent un grain de taille plus petite que la distance effective entre deux grains consécutifs, (disons un grain de 10ms pour une densité de 10Hz) l'oreille va les séparer, et la perception auditive entendra un rythme. Mais, si on fait diminuer la distance temporelle entre ces grains (disons jusqu'à une densité de 500Hz), ils commenceront à se superposer, et la perception auditive commencera à les fusionner : on écoute alors une trame - on passe alors du rythme à la trame. Les trames sonores peuvent encore être agglutinées, superposées, façonnées de telle manière que l'oreille perçoive des masses sonores de texture plus ou moins rugueuse et non plus des grains individuels.

Dans des techniques basées sur la STFT, ainsi que dans des techniques de traitement de la voix comme

PSOLA, la manipulation de l'axe temporel donne au signal une élasticité qui permet l'allongement et le rétrécissement de ce signal sans pourtant perdre ses rapports structuraux internes. Cette possibilité est essentielle dans le cas du traitement de la parole. Ainsi, il est possible d'allonger ou rétrécir une phrase du discours vocal sans perdre son rythme, et donc son intelligibilité,

4.2. Spectre : timbre

La manipulation du signal en temps réel réalisée en utilisant les méthodes STFT, est principalement dirigée vers le spectre du son, conduisant musicalement à une transformation au niveau du timbre. Évidemment, la détermination du timbre du signal sonore ne dépend pas seulement de son spectre, mais celui-ci est néanmoins déterminant. En fait, la manipulation discrétisée des paramètres du signal comme le contrôle du fenêtrage et la définition de l'enveloppe dynamique du son sont aussi des variables qui aident à réaliser des effets de timbre.

Le contrôle discrétisé des paramètres spectraux et temporels permet une manipulation aisée des aspects musicaux qui déterminent le timbre.

En ce qui concerne les méthodes de AGS, la manipulation du timbre est moins directement contrôlable, car on n'a pas l'accès direct aux composants spectraux du son. Malgré cela, l'engendrement de trames de masses sonores de textures variables peut générer perceptiblement des sensations de transformation du timbre. Également, dans des techniques de type « pitch synchrone » comme PSOLA on trouve des phénomènes perceptifs découlant de la manipulation indirecte du timbre.

4.3. Fréquence \square hauteur

Dans le cas des techniques granulaires, la sensation perceptive de hauteur dépend généralement des paramètres de manipulation de la durée et de la densité individuelle des grains, en rapport avec le degré de dispersion dans la re-synthèse du signal.

En ce qui concerne le traitement de la hauteur, la méthode PSOLA devient très efficace. La décomposition du signal en formes d'onde élémentaires, permet un traitement fréquentiel que doit être indépendant de la structure temporel du signal d'entrée.

Un phénomène semblable est engendré au moment de la transposition du signal : la possibilité de changer la hauteur perçue en manipulant uniquement les parties périodiques du signal sans nuire à la structure du signal d'origine. Ce traitement de la fréquence peut impliquer un rétrécissement de ces formes d'onde élémentaires, en additionnant des copies de l'onde sonore élémentaire dans les cas où la transposition se fait vers l'aigu, pour remplir une certaine durée et maintenir la structure du signal, la fluidité et la naturalité du discours dans le cas de la voix, ou bien par un allongement ces formes d'onde élémentaires (en enlevant quelques-unes si nécessaire) quand la transposition se fait vers le grave.

Dans des méthodes du type AGS, l'aspect fréquentiel est aussi manipulable. Par exemple, pour des grains de durée suffisante et de basse densité, l'oreille percevra une transposition fréquentielle appliquée aux grains individuels ; cela engendre une certaine « mélodicité ». Par contre, si les grains sont petits et la densité forte, c'est-à-dire, si nous sommes face à une trame sonore, l'oreille percevra la hauteur globale d'un signal complexe. L'application d'une variable de dispersion à la manipulation fréquentielle des « atomes » d'une trame sonore complexe, engendre une bande fréquentielle et permet la manipulation de la largeur de cette bande.

5. GABOR : MUSICALITÉ EN TEMPS RÉEL

Une manipulation soigneuse des variables impliquées dans les méthodes énoncées plus haut, peut permettre une articulation non perceptible auditivement entre les techniques énoncées. Désormais, le compositeur peut, même en temps réel, *se promener* d'une technique à l'autre sans que l'oreille puisse saisir le moment exact de permutation entre les techniques.

La librairie Gabor, intégré dans l'environnement graphique Max / MSP, explore des possibilités de l'analyse / re-synthèse granulaire au sens général. Cette librairie permet des corrélations entre différentes techniques de façon que la technique devient le simple outil avec lequel le compositeur peut sculpter de sonore. La technique de manipulation devient la « toile de fond » parmi laquelle le compositeur fera sortir la musicalité de son œuvre.

Dans ce sens, nous avons constaté que, à l'aide de la librairie Gabor et en certaines conditions, il est possible l'obtention d'une fluidité perceptive entre les résultats sonores des méthodes de manipulation abordées (STFT, AGS et PSOLA), permettant ainsi une articulation fluide entre eux.

Ainsi, les variables musicales durée-rythme, timbre et hauteur sont corrélés. Dans l'AGS, le traitement de la temporalité peut engendrer perceptivement des transformations sonores de l'ordre du timbre ou même de la hauteur. De même, l'utilisation de la méthode PSOLA engendre, en certaines conditions, des sensations sonores du type granulaire, rythmique, de même que des transformations au niveau du timbre. Pareillement, les techniques basées sur les méthodes STFT, engendrent non seulement des transformations au niveau du timbre, mais peuvent aussi engendrer des changements soit au niveau de la perception d'une texture, soit au niveau du caractère plus ou moins rythmique, soit des simples changements de la hauteur.

6. CONCLUSION

L'utilisation de la librairie GABOR rend la manipulation du signal en temps réel très flexible. Cette librairie permet une grande fluidité d'interaction entre des méthodes diverses de traitement de signal. Ces méthodes peuvent toujours être intégrées de façon telle que le résultat sonore soit celui musicalement souhaité.

7. RÉFÉRENCES

- [1] GABOR, Dennis. (1946). « Theory of communication » in *Journal of the Institut of Electrical Engineers Part III*, 93, p. 429-457.
- [2] GABOR, Dennis. (1947). « Acoustical Quanta and the Theory of Hearing » in *Nature*, 159 (4044): p. 591 – 594. Londre.
- [3] GABOR, Dennis. (1952). « Lectures on communication theory » Technical Rapport 238, Research Laboratory of Electronics. Cambridge.
- [4] PEETERS, Geoffroy. (1998). « Analyse et synthèse des sons musicaux par la méthode PSOLA » in *Actes JIM98*, Angoulême, France.
- [5] PEETERS, Geoffroy. RODET, Xavier. (1999). « SINOLA : A new Analysis / Synthesis Method using Spectrum Peak Shape Distortion, Phase and Reassigned Spectrum » in. *ICMC*.
- [6] ROADS, Curtis. (1996). *The Computer Music Tutorial*. MIT Press.
- [7] ROADS, Curtis. (2001). *Microsound*. MIT Press.
- [8] SCHNELL, N., BORGHESI, R., SCHWARZ, D., BELVILACQUA, F., MÜLLER, R., (2005). « FTM – Complex Data Structure for MAX », (à paraître).
- [9] SCHNELL, N., PEETERS, G., LEMUTON, S., MANOURY, P., RODET, X., (2000). « Synthesizing a choir in real-time using Pitch Synchronous OverLap Add (PSOLA) » in *ICMC*. Berlin.
- [10] SCHNELL, N., SCHWARZ, D., (2005). « GABOR, Multi-representation Real-Time analysis / Synthesis », (à paraître).
- [11] TRUAX, Barry. (1988). « Real time granular synthesis with a digital signal processing computer » in *Computer Music Journal*, vol.12 n°2.