

REFLEXIONS SUR LES RELATIONS MUSIQUE/VIDEO ET STRATEGIES DE MAPPING POUR MAXMSP/JITTER

Benoît Courribet

CICM - Université Paris VIII

MSH Paris Nord

bcourribet@mshparisnord.org

RÉSUMÉ

L'article présente les réflexions de l'auteur concernant les rapports entre musique et vidéo, dans le cadre de la composition d'œuvres musicales hybrides, incluant une composante visuelle. Après quelques rappels et généralités sur les spécificités des media concernés, des stratégies de *mapping* sont présentées et des exemples concrets évoqués. Enfin, à l'issue de l'exposé, les travaux en cours de l'auteur sur une pièce audiovisuelle sont présentés.

1. INTRODUCTION

Dans le cadre de l'étude des rapports entre musique et vidéo, de nombreuses recherches sont en cours au CICM et dans le cadre du groupe de travail de l'AFIM sur la visualisation du sonore.

De manière plus personnelle, j'ai toujours été intéressé par l'impact du médium visuel sur notre perception. De plus, les travaux cinématographiques d'artistes tels qu'Andrei Tarkovsky, David Lynch, Norman McLaren, Lazlo Moholy-Nagy, m'ont donné envie d'intégrer un travail visuel à mon activité de composition musicale.

D'un point de vue technologique, cela est devenu possible grâce à des développements récents: la possibilité de créer des DVD-Vidéos sur un ordinateur personnel, et l'ajout de la bibliothèque Jitter à l'environnement de programmation graphique Max.

Au cours de cette démarche, je me suis constamment positionné comme musicien en essayant de garder la question suivante à l'esprit: Comment ces deux média peuvent-ils co-exister ou interagir étant donné leurs différences et spécificités ontologiques?

Dans cet article, je m'attacherai à présenter quelques éléments de réflexion à propos de cette question, puis proposerai quelques stratégies de mapping entre vidéo et musique. Finalement, un travail en cours concernant une pièce audiovisuelle sera présenté.

2. REFLEXIONS SUR LES RELATIONS MUSIQUE/VIDEO

2.1. Objectifs et orientations

Le but que j'essaie d'atteindre est la réalisation d'une musique ayant une composante visuelle, j'aspire à ce que la vidéo fasse partie de la musique, et non pas qu'elle en soit une illustration ou une représentation, au sens d'une transcription graphique [1]. Cette idée s'est

imposée au cours du visionnage d'œuvres d'Andrei Tarkovsky, qui me sont apparues extrêmement musicales: son travail sur l'articulation dans le temps de structures et éclairages est une excellente illustration de ce que l'on pourrait appeler une "musique pour les yeux". On peut prendre l'exemple d'une des scènes finales dans le film *Stalker* lorsque, après de subtiles variations d'éclairages, la pluie tombe sur une étendue d'eau et illumine la scène toute entière.

De nombreux outils de visualisation des sons existent et sont implémentés dans les lecteurs multimédia. Je les appellerai outils globaux de visualisation car ils consistent en une représentation visuelle des données audionumériques, de la globalité de l'information sonore. La forme d'onde du son ou sa représentation fréquentielle sont utilisées comme matériau de base pour la vidéo sur lesquels des filtres sont appliqués (déformations géométriques...).

Ces outils semblent présenter deux inconvénients majeurs. Le premier est du au caractère orthothétique (qui "pose exactement") [2] de ces représentations: comme elles présentent exactement le phénomène sonore, elles n'y apportent rien, sinon à des fins d'analyse ou de documentation. Le second inconvénient tient à la contextualisation des matériaux sonores et visuels: avec de telles méthodes, les mêmes sons, placés dans des contextes musicaux différents seraient représentés à l'identique. L'effet Koulechov, dans le domaine du cinéma illustre le fait qu'un plan peut être interprété par les spectateurs de manière très différente en fonction des plans qui le précèdent et lui succèdent. Bien entendu, on peut trouver un écho à cet effet dans le domaine musical. Cela met en lumière les mécanismes de rétention et protention décrits par le philosophe E. Husserl [3]. De même, il semble intéressant artistiquement de tirer profit de la possibilité de changer la perception d'un son, lorsqu'il est accompagné d'informations visuelles différentes.

A des outils globaux, je préférerais donc l'idée de stratégies locales pour les interactions entre musique et vidéo, car chaque œuvre musicale est unique et requiert une attention spécifique. Aussi, il semble primordial que les parties sonores et visuelles doivent être composées ensemble afin de tirer parti au maximum de leurs interactions et de contourner le problème d'une réalisation chronologique "musique, puis vidéo". Il me semble essentiel que des liens formels régissant les interactions n'en viennent pas à surdéterminer l'esthétique de l'ensemble, et laissent au compositeur le soin et la possibilité de faire les choix qui lui semblent appropriés sur les plans esthétique et compositionnel.

2.2. Quelques remarques au sujet de spécificités ontologiques de la musique et de la vidéo.

Tout d'abord, si l'on s'intéresse aux relations entre ces média et l'espace, on constate que la distribution et la propagation du son dans l'espace est tridimensionnelle, alors que la vidéo est la plupart du temps projetée ou affichée sur un écran rectangulaire bidimensionnel, ce qui implique une posture physique orientée face à l'écran pour la réception des informations visuelles. De plus, des sons placés en arrière-plan et en avant-plan peuvent coexister dans l'espace d'écoute sans pour autant se cacher mutuellement, contrairement à ce qui se passe dans le domaine de l'image. Ceci laisse présager des difficultés à présenter à l'écran une contrepartie aux attributs spatiaux de la musique ou des matériaux sonores.

En ce qui concerne le problème du temps, les différences sont plus marquées encore. Les sons n'existent pas en dehors du temps; les sons sont, par définition, dus à des variations de pression de l'air au cours du temps. Au contraire, même si notre perception visuelle s'inscrit dans un flux temporel, un arrêt dans un flux vidéo, une image fixe, porte toujours une signification visuelle. De plus, les matériaux sonores et visuels ne se comportent pas de la même manière lorsqu'on leur applique des distorsions temporelles. Nous reconnaitrions toujours la même chose si une vidéo était lue à une vitesse dix fois inférieure à la vitesse de lecture originale, alors qu'un son serait méconnaissable. Pierre Schaeffer, à propos de la réalisation de son "Etude aux chemins de fer" disait qu'à l'opération quantitative du ralentissement d'un son de train de 78 à 33 tours était liée une modification qualitative sur la nature du son lui-même [4].

On peut également évoquer le caractère discret de la vidéo, qui fait qu'il est difficile de ralentir un flux vidéo, car les paramètres de discrétisation (c'est à dire le nombre d'images par seconde) sont basés sur et sont proches des limites de notre perception de la continuité visuelle.

Enfin, nous sommes familiers avec la notion de non-linéarité dans la perception auditive humaine (qu'il s'agisse de notre perception des intensités ou des fréquences) et, bien évidemment, c'est également le cas pour notre perception des intensités de couleur et lumière. Le fait que l'on ait à travailler avec les deux à la fois montre l'importance du choix des techniques de mapping entre musique et vidéo.

3. STRATEGIES DE MAPPING MUSIQUE/VIDEO AVEC JITTER

Considérant les travaux effectués dans le domaine de la visualisation du son au CICM [5], quelques exemples de stratégies de mapping musique/vidéo vont être exposés selon deux approches principales: une approche séquentielle et une approche synchrone.

3.1. Approche séquentielle

J'appelle une approche séquentielle le fait d'extraire des données du son pour piloter des paramètres de synthèse ou traitement vidéo. Malgré cette chronologie, cette approche n'est pas en contradiction avec les considérations précédentes, car elle peut être appliquée à

des matériaux sonores et donc intervenir avant le processus de composition.

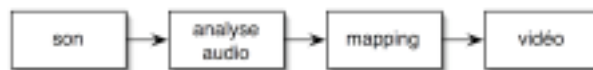


Figure 1. Schéma résumant l'approche séquentielle

Cette approche peut être considérée comme une continuation des travaux de Vincent Verfaillie [6], mais avec des media croisés. En effet, Vincent Verfaillie a consacré sa thèse de doctorat aux effets audio adaptatifs : il s'agit de contrôler dynamiquement les paramètres de traitements audio numériques à l'aide de paramètres extraits du son d'origine. Il insiste sur l'importance du mapping dans ce processus.

De nombreux outils ont été développés afin d'extraire des paramètres perceptifs des sons, citons par exemple l'objet *analyze~* de Tristan Jehan¹. Cet objet permet, à l'aide d'une analyse spectrale, d'obtenir des informations comme les paramètres de *loudness*, *noisyness*, *brightness* (centre de gravité spectrale), détection d'attaque... Afin de permettre une plus grande flexibilité, un patch Max/MSP a été réalisé afin d'enregistrer sous forme de pistes audio ces paramètres. On obtient ainsi un fichier audio multipiste dans lequel les paramètres extraits sont synchronisés avec le son d'origine.

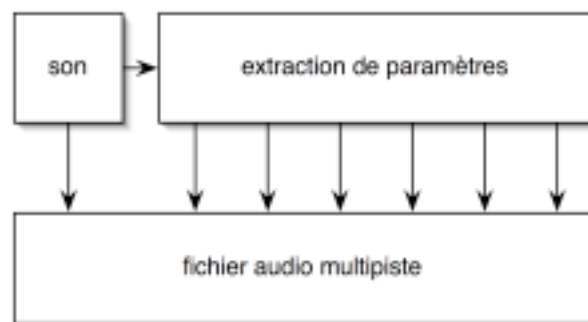


Figure 2. Enregistrement de paramètres sous forme de données audio.

On peut ainsi obtenir un fichier 8 pistes dans lequel les 2 premières pistes sont le fichier audio stéréo lui-même, les pistes suivantes stockant pour chacun des 2 canaux le centre de gravité spectrale et les paramètres de *loudness* et *noisyness*.

Ce fichier multipiste est ensuite utilisé comme source pour modifier un fichier vidéo. Les différentes pistes, après passage par des fonctions de transfert sont affectées à des paramètres de traitement vidéo.

¹ <http://web.media.mit.edu/~tristan/>

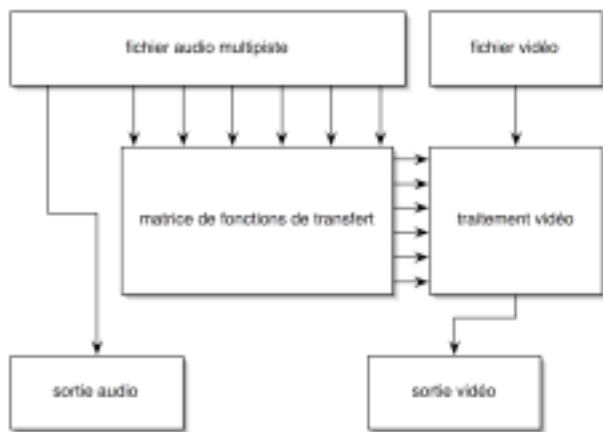


Figure 3. Schéma de la mise en œuvre du mapping audio/vidéo dans le cas d'une approche séquentielle.

Par exemple, le paramètre de *loudness* des canaux gauche et droite est extrait afin de piloter le ratio de zooms gauche et droit d'une vidéo après un ajustement à l'aide d'une fonction de transfert. En fonction du choix du matériau vidéo, de la fonction de transfert et d'un filtre approprié pour lisser le résultat vidéo final, le résultat peut convenir à de nombreux types de sons.

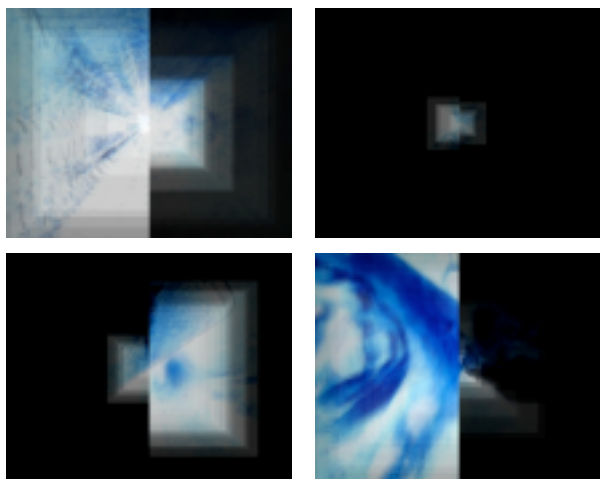


Figure 4. Extraits de matériaux vidéo pour la pièce "time.mass.velocity.light".

Sur un autre exemple, des attaques sont détectées afin de provoquer un changement de configuration d'une structure 3D aplatie: la scène 3D est effacée lorsqu'une attaque est détectée, sinon, les images successives se superposent pour créer des formes 2D abstraites. L'idée visuelle est un écho aux travaux de l'artiste Lazlo Moholy-Nagy et notamment la vidéo "Lichtspiel: Schwartz-Weiss-Grau" dans laquelle, c'est le jeu d'ombres et de lumières provenant d'une structure mobile qui est filmé. Le mouvement de formes géométriques abstraites qui en résulte est régi par les lois physiques et donne une cohérence structurelle au résultat projeté.

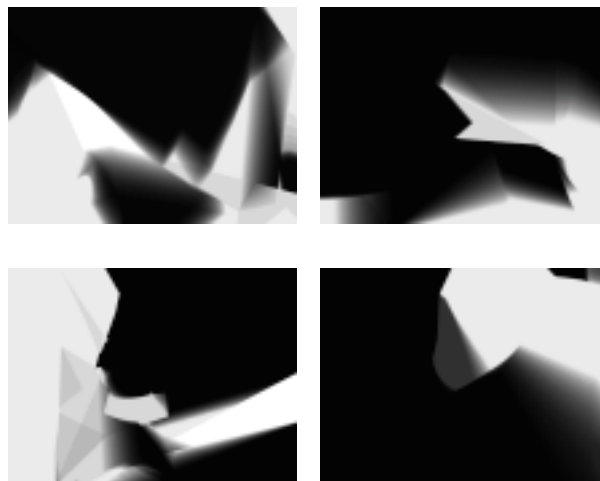


Figure 5. Extraits de matériaux vidéo pour la pièce "time.mass.velocity.light".

Ceci a été utilisé sur des sons de rebonds et sur de la percussion temps réel.

3.2. Approche synchrone

L'approche synchrone consiste à faire contrôler les parties audio et vidéo par un même jeu de paramètres dynamiques (que nous appelons paramètres intentionnels car ils pré-existent à l'audio et à la vidéo). L'avantage de cette approche est que parfois certaines données ne peuvent être extraites d'un son ayant subi un traitement. Par exemple, il est dur d'extraire un temps de délai d'un son traité par une ligne à retard avec réinjection.



Figure 6. Schéma résumant l'approche synchrone

Cette approche a été utilisée à la fois dans des contextes temps réel et temps différé. La pièce Isol contenait des parties audio et vidéo, reliées par le même moteur de granulation temps réel: un percussionniste jouait une partie de cymbale semi-improvisée; un microphone de contact placé sur la cymbale était connecté à un ordinateur afin de rendre possible un traitement temps réel du son entrant. Une horloge aléatoire était utilisée pour générer des grains de sons et en même temps déformer une structure 3D et changer la position de la caméra dans la scène 3D. La pièce consistait en des changements continus des morphologies sonores et visuelles, les paramètres de réinjection audio et de rémanence vidéo, tailles de grains et taille des structures étant liés.

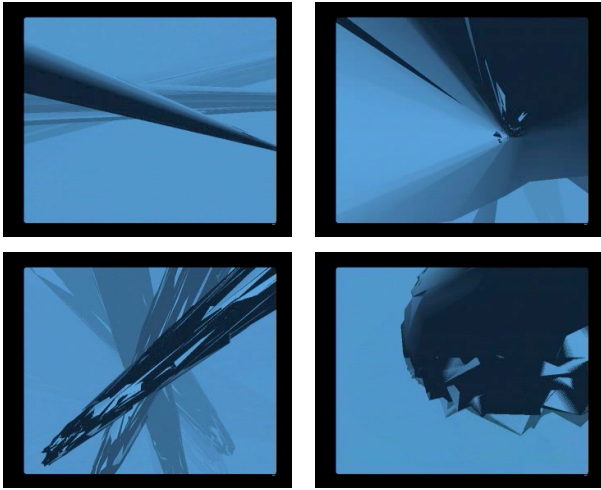


Figure 7. Extraits de la vidéo de la pièce "isol".

Dans un contexte temps différé, la même horloge fut utilisée pour piloter un moteur de granulation audio et pour provoquer des discontinuités dans la lecture d'une vidéo. En fonction de la taille des grains et de leur densité, de nombreux effets visuels pouvaient être obtenus, allant de subtils brouillages temporels à des segmentations aléatoires extrêmes. Le choix du matériau vidéo est extrêmement important, et de lents zooms ou travellings sont des matériaux vidéos donnant d'excellents résultats.

4. TIME.MASS.VELOCITY.LIGHT: WORK IN PROGRESS.

Au cours des six derniers mois, mon travail de composition s'est concentré sur une pièce de musique électroacoustique, "time.mass.velocity". Elle est maintenant achevée et sert à présent de base à la réalisation d'une pièce audiovisuelle. L'idée générale n'est pas de préparer une piste vidéo pour l'accompagner, mais plutôt de composer une pièce incluant de la vidéo, se basant sur les mêmes matériaux sonores que la pièce originale ainsi que certaines de ses articulations. L'avantage de préparer cette pièce pour un support DVD-Vidéo réside dans la possibilité d'utiliser également le format audio 5.1 pour la spatialisation des sources audio.

La méthode de composition s'inscrit dans la tradition de la musique concrète: le concept, la forme et les articulations intrinsèques de l'oeuvre émergent de la manipulation concrète des sons. Mon travail est également influencé dans une large mesure par le compositeur Horacio Vaggione, par ses réflexions et méthodes pour la composition électroacoustique, en particulier, l'utilisation de transformations itératives et la notion de saillances, décrites dans de nombreux articles [7]. Ces principes permettent d'atteindre un haut niveau de complexité (au sens de Lévy-Leblond [8]) en maintenant une certaine homogénéité parmi les matériaux sonores, étant donné qu'ils dérivent tous d'une même série de sons.

Etant donné que la pièce originale est articulée suivant des concepts relatifs à notre perception et à nos représentations mentales du temps ou de ses attributs, une série d'outils logiciels a été réalisée pour proposer

une large palette de traitements basés sur un paradigme temporel (micro-montage automatique, granulation, distorsions/délinéarisations temporelles, décorrélations multiples et variables...).

Je suis donc en train de développer une autre série d'outils de traitement vidéo basés sur les mêmes concepts, dans le but de, à l'aide des matériaux sonores d'origine, de créer et modifier les matériaux vidéo de manière synchrone au traitement des sons. Je suis principalement confronté à une difficulté majeure à propos de la vidéo: je voudrais être capable de sculpter le temps du matériau de la même manière que je le fais pour le son, mais cela nécessiterait un équipement d'acquisition vidéo spécial (une caméra haute-vitesse), afin de réaliser des ralentissements extrêmes sans modifier l'impression de continuité du flux vidéo. Lorsque le développement logiciel sera achevé et les matériaux audio et vidéo créés et traités, le processus de composition commencera, il faudra agencer dans l'espace et dans le temps à la fois l'audio et la vidéo pour réaliser la pièce.

5. CONCLUSION

Des réflexions, stratégies de mapping et exemples d'utilisation des relations entre musique et vidéo ont été présentés.

Il ressort de ces réflexions qu'une attention toute particulière doit être accordée à la méthodologie de la mise en oeuvre de ces techniques, afin que musique et vidéo ne soient pas considérées comme indépendantes ou dans des rapports d'illustration ou de représentation. En effet, il semble primordial que la musique et la vidéo soient dans une relation de transduction [9], au sens où elles se constituent mutuellement dans leur rapport l'une envers l'autre; une relation qui a "valeur d'être". On assiste alors, pour reprendre le vocabulaire de G.Simondon, au cours du travail de composition, à l'individuation d'une oeuvre hybride. Cette oeuvre appartient s'apparente à ce qu'on pourrait appeler un « art multimédia du temps » car, pour la partie sonore autant que pour la partie visuelle, c'est l'articulation des matériaux dans le temps, leur mise en forme inscrite dans le temps qui m'intéresse tout particulièrement.

6. REFERENCES

- [1] Régnauld, C., "Correspondances entre graphisme et son: les représentations visuelles de l'objet sonore" in *Du sonore au musical : cinquante années de recherches concrètes (1948-1998)* Sous la direction de Sylvie Dallet et Anne Veitl, l'Harmattan, Paris, France, 2001.
- [2] Stiegler, B. *La technique et le temps ; Tome 2 : la désorientation* Galilée, Paris, France, 1996.
- [3] Husserl, E. *Leçons pour une phénoménologie de la conscience intime du temps*. PUF, Paris, France, 1964.
- [4] Von der Weid, J.N. *La musique du 20^{ème} siècle*, p.213, Hachette, Paris, France, 1997.

- [5] Sedes, A., Courribet, B. et Thiébaud, J.B., " From the visualization of sound to real-time sonification: different prototypes in the Max/MSP/Jitter environment." *Proceedings of the 2004 International music Conference*. International Computer Music Association, Miami, USA, 2004.
- [6] Verfaillie V., *Effets Audionumériques Adaptatifs: Théorie, Mise en Oeuvre et Applications en Création Musicale Numérique*, mémoire de thèse de doctorat, Université Aix-Marseille, Marseille, France 1988.
- [7] Vaggione, H., "Vers une approche transformationnelle en CAO" *Actes des JIM 1996*, Les cahiers du GREYC, CNRS-Université de Caen, Caen, France, 1996.
- [8] Lévy-Leblond, J.M., "La physique, une science sans complexe?" *Les théories de la complexité*, Seuil, Paris, France, 1991.
- [9] Simondon, G., *L'individu et sa genèse physico-biologique*, Presses Universitaires de France, Paris, France, 1964.