

Plumage: Design d'une interface 3D pour le parcours d'échantillons sonores granularisés

Christian Jacquemin et Rami Ajaj

Roland Cahen et Yoan Olivier

Diemo Schwarz

LIMSI-CNRS & Univ. Paris 11
91403 Orsay, France
Prenom.Nom@limsi.fr

ENSCI
75011 Paris, France
Prenom.Nom@ensci.fr

IRCAM
75004 Paris, France
Prenom.Nom@ircam.fr

RESUME

Plumage est une interface pour le contrôle interactif de la composition et du jeu de scènes graphiques et audio spatialisées. L'interface repose sur un environnement graphique 3D composé de têtes de lecture animées parcourant un espace sonore matérialisé par des plumes associées à des grains sonores (des micro-échantillons). La disposition spatiale des plumes repose sur des paramètres sonores des échantillons qu'elles représentent. Le jeu musical est une lecture interactive et continue de cet espace sonore contrôlée par manipulation directe des paramètres des trajectoires des têtes de lecture. *Plumage* porte sur le design d'interaction. Il rend tangibles, organise et synchronise des lectures d'échantillons avec leurs comportements graphiques et sonores pour une navigation audio-graphique efficace.

MOTS CLES : Multimodalité, Composition audio-visuelle, Interfaces 3D.

ABSTRACT

Plumage is an interface for interactive 3D audio/graphic scene browsing and design. The interface relies on the notion of tape heads in a sonic and graphic 3D space made of feathers associated with sound micro-samples. The spatial layout of the feathers is defined by sound parameters of the associated samples. The musical play is the outcome of a continuous and interactive navigation in the sound space controlled by direct manipulation of tape head trajectories. This work is based on the design of interactions. A simple and efficient audio graphic navigation is designed through the combination and synchronization of elementer sound object triggerings.

CATEGORIES AND SUBJECT DESCRIPTORS: H5.m. Information interfaces and presentation (e.g., HCI) : Miscellaneous.

GENERAL TERMS: Documentation, Instructions

KEYWORDS: Multimodality, Audio-visual composition, 3D GUIs.

INTRODUCTION

Le rendu audio-visuel spatialisé est au cœur des problématiques de la réalité virtuelle et augmentée lorsqu'on ne souhaite pas se limiter à une seule modalité (la vue). Afin d'offrir un rendu cohérent, le positionnement spatial des sources sonores et celui des objets géométriques doivent être définis dans un même référentiel et une même temporalité. Des formalismes tels que MPEG-4 BIFS [1] ou X3D¹ peuvent recevoir ce type de description multimédia.

Ces formalismes et les applications qui proposent des compositions audio-visuelles spatialisées reposent sur une description de sources ponctuelles et leur positionnement dans l'espace. A contrario, les techniques de rendus d'image pour une approche continue des propriétés graphiques de l'espace (comme la fumée ou le brouillard) utilisent une représentation spatiale discrétisée des propriétés lumineuse du média [2]. Notre travail offre une première convergence entre une représentation spatiale continue du son et des propriétés lumineuses. Il décrit un nouveau type d'interface audio-graphique spatialisé dans laquelle le son est représenté par une distribution spatiale de grains sonores et le rendu sonore se fait par parcours de lecture spatiale de ces grains (figure 1). Ce travail ouvre de nouvelles perspectives sur le rendu audio-graphique spatial non restreint à des sources ponctuelles.

L'interaction est au cœur de cette interface intitulée *Plumage* et composée de petits éléments visuels en forme de plumes, placés dans l'espace géométrique en fonction de paramètres de l'espace sonore du grain qu'ils représentent. La navigation interactive vise à s'en approcher pour les écouter et se construire un parcours multi-sensoriel perceptuel dans l'univers composé d'ensembles de grains/plumes. Ces objets audio-graphiques sont distribués et représentés visuellement selon un mapping de descripteurs sonores sur les paramètres graphiques. Le parcours de navigation sonore s'appuie sur des modalités de navigation audio-graphiques diverses et originales : parcours évolutifs de têtes de lecture multiples, positions de caméras et de micros dissociés...

¹<http://www.web3d.org/>



FIG. 1 : Design graphique de *Plumage*.

Plumage résulte de la coopération interdisciplinaire entre designers et chercheurs dans le cadre du projet *Énigmes*² piloté par Roland Cahen (ENSCI) et portant sur les « partitions navigables ». Il reprend des hypothèses et des recherches conduites dans le cadre du projet collaboratif *Phase* [4] sur l'utilisation d'un environnement virtuel 3D et d'un bras à retour d'effort pour parcourir un espace sonore interactif. Par partitions navigables, l'on entend ici des formes de notation ou de représentation musicale actives, avec lesquelles on peut directement jouer la musique. L'outil numérique permet de ne pas dissocier la partition de l'instrument et propose, par la notion de navigation sonore, de nouvelles approches de la composition/interprétation de la musique et du son.

Il ne s'agit ni de naviguer dans des bases de données car les corpus sonores utilisés sont réduits à la segmentation d'une ou quelques séquences sonores, ni d'analyser des œuvres musicales car nous restons, pour la matière sonore, à un niveau d'unités temporelles quasi corpusculaires, ni même de visiter avec un rendu l'espace perceptif des timbres d'une base de sons d'instruments³, mais de travailler sur le design de la navigation audio-graphique liée à ces objets. Si c'est une réalité virtuelle ou une virtualité elle est ici abstraite, bien qu'elle utilise les modalités spatiales communes de la navigation.

ARCHITECTURE

L'architecture générale de *Plumage* est décrite dans la figure 2. Elle est organisée autour de deux applications complémentaires. CATART [6] est un logiciel de synthèse sonore concaténative qui s'appuie sur des grains sonores (des micro-échantillons) garnis de méta-données issus soit de l'analyse des sons, soit de descripteurs de haut niveau.

*Virtual Choreographer*⁴ (VIRCHOR) est un moteur de rendu

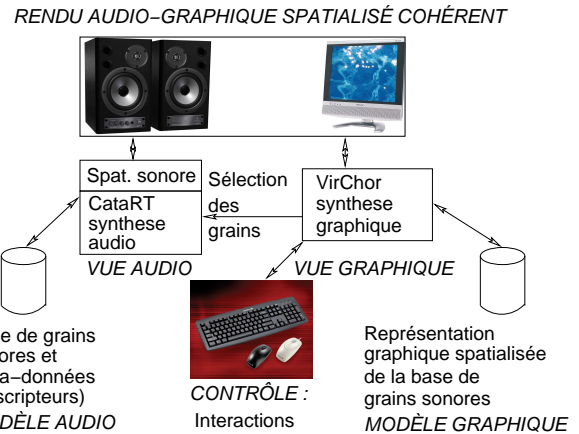


FIG. 2 : Architecture de *Plumage*.

3D temps réel interactif. Deux fonctionnalités de VIRCHOR sont exploitées dans le cadre de *Plumage*. D'une part, des capteurs sont placés dans la scène afin d'émettre des messages vers les objets sonores qu'ils intersectent et en déclencher le jeu. D'autre part, des micros sont positionnés afin de reporter vers un spatialiseur sonore en sortie de CATART les coordonnées sphériques des sources sonores actives dans le référentiel des micros.

INTERFACE DE RENDU AUDIO-VISUEL INTERACTIF

Plumage est une application de rendus graphique et audio cohérents, temps réel et interactifs. L'espace sonore est constitué de grains sonores avec leurs descripteurs. L'espace graphique se compose d'un ensemble de plumes, chaque plume est associée à un grain sonore. Son positionnement spatial et sa représentation graphique sont fonction des méta-données associées aux échantillons sonores (voir figure 3). Au sein de la distribution spatiale des plumes se déplacent trois têtes de lecture sur des trajectoires elliptiques. Ces têtes sont composées d'un micro autour duquel tournent trois capteurs comme des satellites autour d'une planète.

Les plumes sont des objets géométriques avec des propriétés sonores (un échantillon sonore) et un comportement (un script définissant les actions réalisées lorsqu'un événement est reçu). Les têtes de lecture parcourent à vitesse contrôlable leurs trajectoires et sont munies de capteurs. Lorsqu'un de ces capteurs entre dans une plume, elle reçoit un message et déclenche la lecture de l'échantillon sonore associé.

Les micros actifs sont utilisés pour reporter à CATART les identifiants des sources sonores et leurs positions relatives. À partir de ces informations, CATART produit une sortie sonore continue par synthèse granulaire en combinant les informations audio des échantillons associés aux sources actives. À partir des informations sur la position relative des sources par rapport aux micros allumés, le spatialiseur produit une sortie audio cohérente avec le déplacement des têtes dans la scène graphique. Les têtes de lecture sont mues par un système d'automatisation procédurale. Chacune ayant une certaine autonomie dont l'utilisateur contrôle le comportement.

²<http://projetenigmes.free.fr/>

³Il existe des points et des débouchés communs avec des démarches telles que celles de Roel Vertegaal [7] ou Stephen E. McAdams et Bennett Smith [3] dans les domaines de la sonification, du data mining...

⁴<http://virchor.sf.net>

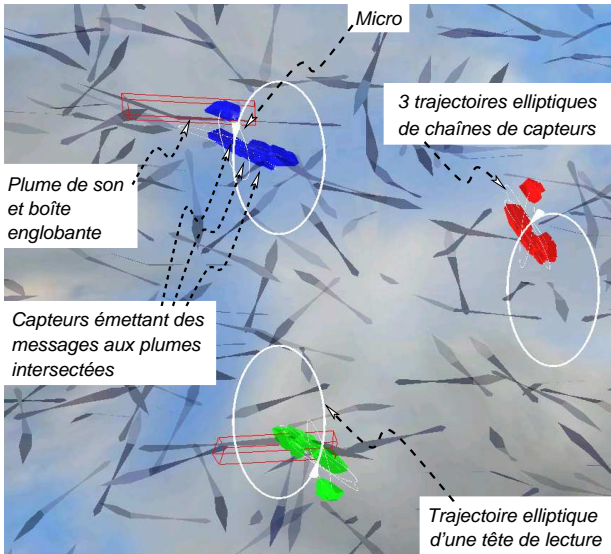


FIG. 3 : Vue d'ensemble de l'interface.

ANALYSE ET RESYNTHESE SONORE

La partie audio du dispositif Plumage est assurée par le système CATART [6] de synthèse concaténative par corpus en temps réel⁵. Ce nouveau paradigme de synthèse sonore musicale utilise une base de données de sons et un algorithme de *sélection d'unités* qui choisit les segments des sons qui conviennent le mieux pour la séquence musicale que l'on souhaite synthétiser, dite la *cible* [5]. La transposition de ce principe à une utilisation en temps réel peut aussi être vue comme une extension basée sur le contenu de la synthèse granulaire, où le seul mode d'accès par position temporelle dans un fichier son est augmenté par un accès ciblé par caractéristiques sonores dans un ensemble de sons.

Le logiciel CATART est implémenté dans MAX/MSP⁶ avec les extensions *FTM* et *Gabor*⁷. Dans une phase d'initialisation, il découpe un ou plusieurs enregistrements en petites unités appelées *grains*, et les caractérise par leurs contenus sonores. Ces paramètres sont donnés par la valeur moyenne de 17 descripteurs sonores tels que la fréquence fondamentale, le volume sonore, la brillance, le niveau de bruit, la répartition spectrale et d'autres caractéristiques du timbre. Les valeurs décrivant des informations de découpage sont également stockées comme descripteurs : la position, la durée et la provenance du segment dans les fichiers son d'origine.

CATART est utilisé dans des contextes musicaux de composition et de performance variés en utilisant une interface simple représentant une projection 2D de l'espace de descripteurs, et une navigation avec la souris, où les grains sont sélectionnés et joués par proximité géométrique. On s'aperçoit très vite des limitations de cette interface, qui permet de visualiser 3 descripteurs seulement (2 axes et couleur des points) et de naviguer selon 2 descripteurs.

⁵<http://imtr.ircam.fr/index.php/CatART>

⁶<http://www.cycling74.com>

⁷<http://ftm.ircam.fr>

Descripteur Sonore	Paramètre Graphique
spectral centroid	x
loudness	y
periodicity	z
pitch	hue
spectral tilt	rotation

TAB. 1 : Exemple de correspondances entre descripteurs et paramètres graphiques.

Dans *Plumage*, CATART échange des données avec VIRCHOR, qui contrôle la sélection des grains (voir la section *INTERACTIONS* ci-dessous). La synthèse dans CATART tient compte des *voix* de grains qui peuvent sonner sur une certaine durée et en parallèle : durée du jeu (une fois, en boucle, ou en continu), multiplicité des échantillons activés par les têtes de lectures et multiplicité des points d'écoute.

MISE EN ŒUVRE

Plumage a été implémenté en mettant en commun des compétences en design d'interface (l'ENSCI), en analyse et resynthèse sonore (l'IRCAM), et en réalisation d'interfaces immersives 3D interactives (le LIMSI-CNRS). Après une phase de prototypage dans *Rhino*⁸, un outil de modélisation 3D, l'interface a été développée conjointement dans deux applications parallèles traitant de l'audio et du graphisme, communiquant entre elles par messages réseau utilisant le protocole OSC⁹, et communiquant avec les utilisateurs via des périphériques externes. Le développement de la plate-forme interactive a conduit à limiter les ambitions initiales (en particulier en termes d'édition dynamique de trajectoires), mais a permis en contrepartie d'explorer des pistes qui n'avaient pas été envisagées initialement sur le comportement interactif des plumes, des micros et des capteurs.

INTERACTIONS

Lors du lancement de l'application, CATART envoie à VIRCHOR un ensemble de paramètres spatiaux et graphiques qui permettent de positionner les plumes dans l'espace (translation et rotation) et de décrire leurs paramètres graphiques (couleur et taille). Ces paramètres sont dérivés d'un sous-ensemble des descripteurs sonores, choisi par l'utilisateur. Un exemple de cette association, utilisée pour les présentations de *Plumage* est donné dans le tableau 1.

Les interactions de l'utilisateur définissent la navigation dans l'espace audio-visuel en envoyant à l'application graphique des paramètres d'animation : taille et orientation des trajectoires des têtes de lecture et des capteurs, vitesse de parcours de ces trajectoires, taille des capteurs... Les caméras peuvent également être modifiées dynamiquement, passant de vues subjectives animées sur les trajectoires à une vue externe surplombant la scène. En modifiant ainsi trajectoires, ouvertures et positions des micros et caméras, on contrôle le parcours

⁸<http://www.rhino3d.com/>

⁹<http://opensoundcontrol.org/>

de l'espace géométrique et sonore, et donc l'exploration des combinaisons entre les rendus graphique et audio corrélés.

PERCEPTION DES INTERACTIONS MUSICO-VISUELLES

La navigation sonore dans le dispositif revient à se déplacer dans un espace virtuel dans lequel des objets sonnent lorsqu'une tête de lecture les approche ou les touche. Cette modalité audio-graphique nous rapproche de notre expérience quotidienne du réel en ce qu'elle donne aux objets sonores des comportements physiques ou symboliques situés spatialement, dont la distribution et l'ordre musical variables définissent le parcours de l'utilisateur. La représentation visuelle fait fonction de repérage et informe sur les potentialités physiques sonores des objets rencontrés. Cette ressemblance avec l'expérience du réel tire parti de nos capacités sensorielles à interpréter un nombre important de phénomènes, leur défection, variabilité... La corrélation spatiale et temporelle des événements audio et visuels et la capacité à se mouvoir apportent une grande précision spatiale et temporelle pour percevoir les événements et les objets situés.

Nous avons constaté une certaine rareté des rencontres entre une tête de lecture simple (un point) et les objets sonores (plumes ou feuilles). Après avoir imaginé des curseurs à dimensions variables (point, ligne, surface, volume), nous avons choisi d'intensifier les rencontres en multipliant les têtes de lectures et les micros et en créant un mécanisme permettant à la fois le brassage des grains, les variations autour d'une position et d'une trajectoire et la polyphonie. Les jeux de têtes de lectures multiples offrent une combinatoire de jeux plus riche qui augmente les rencontres avec les objets sonores et produit des sons plus diversifiés. La variation et la capacité de mixage des points d'écoute (équivalents sonores du point de vue ici incarnés par les micros) offrent un contrôle dynamique sur l'écoute : soit elle a lieu depuis la position de la caméra, soit elle est placée sur le curseur ou la tête de lecture.

Nous avons travaillé avec un ensemble de 200 échantillons et nous sommes rendus compte à l'usage que malgré le grand nombre d'éléments sonores, on se familiarise très vite avec la navigation dans l'environnement sonore obtenu et on en perçoit à l'oreille l'exhaustivité fondamentale. On pourrait en avoir 10 ou 100 fois plus à disposition et très probablement s'y retrouver tout aussi facilement. Plus profondément, la question qui se pose au musicien est celle de la déconstruction et de la reconstruction. En effet, le découpage en grains déconstruit le son que la navigation doit reconstruire d'une autre manière. Les structures intrinsèques des extraits sonores ou musicaux sont perdues lors du découpage en grain. Ce qui reste est un classement des timbres composant le corpus sonore d'origine, un peu comme si on rangeait l'ensemble des touches de couleurs composant un tableau dans l'ordre des différentes gammes de couleurs.

L'expérience de cette redondance musicale peut devenir un atout car elle témoigne d'un potentiel à percevoir des variations plus importantes, elle nous encourage à approfondir les recherches et les applications. Se pose alors la question

des usages, difficiles à définir a priori, mais qui déterminent nécessairement les modes d'interaction et les attentes.

Sur le plan des applications possibles voici quelques pistes sur lesquelles nous travaillons. **Artistique** : installation immersive de réalité virtuelle abstraite avec commande gestuelle pour se mouvoir dans une forêt sonore à la manière des sculptures de Soto... **Musique** : contrôle gestuel, rendu et repérage audio-graphique des trajectoires de plusieurs têtes de lecture dans l'interprétation (rejeu) des musiques sur support. **Scientifique** : sonification et data mining, amélioration des rendus audio-graphiques **Technique** : repérage audio-graphique dans l'espace 3D, optimisation des flux de synchronisation et échanges de données audio-graphiques. **Ludoéducatif** : découverte comparative de jeux de timbres sonore, être à l'intérieur d'une musique. **Urbanisme** : cartographie sonorisée interactive.

SYNTHÈSE ET PERSPECTIVES

Le projet *Plumage* a donc permis de mettre en œuvre un espace multimédia audio/graphisme de navigation synchrone, cohérente et contrôlable en temps réel. Il offre de nouvelles perspectives pour la composition musico-graphique et le jeu paramétré et continu de cet espace. De nombreuses questions scientifiques sur la correspondance transmodale dynamique entre son et image restent encore à explorer : la correspondance entre les dérivées temporelles des variables géométriques (vitesse et accélération linéaires et angulaires) et leurs dépendances avec des effets sonores tels que l'effet Doppler, et le lissage des informations spatiales et sonores envoyées au synthétiseur.

BIBLIOGRAPHIE

1. P. Daras, I. Kompatsiaris, T. Raptis, and M. G. Strintzis. MPEG-4 authoring tool for the composition of 3D audio-visual scenes. In *DCV '01*, page 110. IEEE, 2001.
2. H. W. Jensen. *Realistic Image Synthesis Using Photon Mapping*. AK Peters, Wellesley, MA, USA, 2001.
3. S. McAdams, B. K. Smith, S. Vieillard, E. Bigand, and R. Reynolds. Real-time perception of a contemporary musical work in a live concert setting. In *7th Intl. Conf. on Music Perception and Cognition*, 2002.
4. X. Rodet, J.-P. Lambert, R. Cahen, T. Gaudy, F. Gosselin, and F. Guédy. Sound and music control using haptic and visual feedback in the PHASE installation. In *NIME'05*, 2005.
5. D. Schwarz. Corpus-based concatenative synthesis. *IEEE Sig. Proc. Mag.*, 24(1), Mar. 2007.
6. D. Schwarz, G. Beller, B. Verbrugge, and S. Britton. Real-Time Corpus-Based Concatenative Synthesis with CataRT. In *DAFx-06*, Montreal, Canada, Sept. 2006.
7. R. Vertegaal and B. Eaglestone. Looking for sound? : selling perceptual space in hierarchically nested boxes. In *CHI '98*, pages 295–296, New York, 1998. ACM Press.