

LOOPJAM: UNE CARTE MUSICALE COLLABORATIVE SUR LA PISTE DE DANSE

Christian Frisson, Stéphane Dupont, Julien Leroy, Alexis Moinet, Thierry Ravet, Xavier Siebert, Thierry Dutoit
Université de Mons (UMONS), laboratoire TCTS et MathRO
Boulevard Dolez 31 B-7000 Mons, Belgium
Courriels: *prenom.nom@umons.ac.be*

RÉSUMÉ

Ce papier présente l'installation LoopJam qui permet aux visiteurs d'interagir avec une carte musicale par le biais d'un système de suivi gestuel par vision informatique. La carte sonore résulte d'un partitionnement des sons en groupes par similarité basée sur leur signal. Le rendu sonore est contrôlé par les positions ou gestes des participants captés par une caméra Kinect détectant la profondeur de la scène 3D. Les mouvements des participants exprimant une mesure ou un tempo sont corrélés à la vitesse de lecture commune à tous les échantillons synchronisés par le moteur audio. Nous avons présenté et testé une première version de cette installation lors de trois expositions en Belgique, Italie et France. Les réactions parmi les participants ont varié entre la curiosité et l'amusement.

1. INTRODUCTION

L'intention de l'installation LoopJam est d'inciter ses visiteurs à construire collaborativement une atmosphère musicale en direct. Une carte sonore bidimensionnelle est constituée automatiquement par notre logiciel qui analyse les échantillons sonores et les regroupe par similarité. En parcourant l'installation, chaque visiteur explore cette carte sonore et déclenche des événements sonores par des gestes simples qui déterminent également le tempo. La lecture de chaque son est synchronisée par notre logiciel, permettant ainsi une composition collaborative musicale en direct créée par tous les visiteurs simultanés.

Dans la partie 2 nous dégagons le contexte de cette installation et décrivons brièvement les travaux connexes. Dans la partie 3 nous détaillons l'architecture logicielle et les choix d'implémentation. Dans la partie 4 nous discutons des conclusions après avoir observé des visiteurs utilisant l'installation. Dans la partie 5 nous proposons de potentiels améliorations que nous envisageons pour cette installation.

2. TRAVAUX SIMILAIRES

2.1. Domaines d'application

L'objectif principal de l'installation LoopJam est de permettre aux visiteurs de créer un rendu musical interactif et collaboratif. Les ancêtres informatisés les plus di-

rects à cette installation depuis les années 1950 sont les environnements de synthèse sonore, donnant accès à des paramètres d'algorithmes de traitement de signal audio, et les séquenceurs audio multipistes, inspirés directement des techniques tangibles de découpage et recollage de bandes magnétiques. Ces systèmes ont proposé des interfaces utilisateur à convivialité variable, à l'origine à l'aide de cartes trouées ou de terminaux par saisie de lignes de commandes, en temps différé selon les limitations de traitement informatique ; par la suite des interfaces graphiques associées à des souris, claviers et autres contrôleurs munis de potentiomètres rotatifs ou linéaires, etc... fournissant une interactivité plus attirante, en temps-réel. Récemment un intérêt fort a été porté sur les interfaces utilisateur tangibles and les dispositifs multi-point, dans l'intention de regagner une interaction gestuelle avec la musique, pourtant les interfaces "WIMP" (de l'anglais : Fenêtre Icône Menu Pointeur) sont toujours majoritairement employées dans les studios des professionnels de la musique.

L'arrivée des ordinateurs personnels et des tablettes a popularisé l'utilisation de ces systèmes, des experts et chercheurs vers les amateurs. Une convergence entre les milieux des performances musicales et de l'industrie du jeu vidéo, particulièrement les séries *Guitar Hero* et *Rock Band*, a suscité un intérêt pour la création musicale chez davantage d'utilisateurs pas nécessairement virtuoses mais désirant allier créativité et simplicité.

2.2. Une taxonomie

LoopJam combine un ensemble des technologies éprouvées et émergentes : le traitement de signal audio et musical, la fouille dans les données musicales, la navigation audio, la vision informatisée 3D, le contrôle gestuel, la reconnaissance d'expressions corporelles.

Les travaux connexes peuvent être classés selon les caractéristiques suivantes :

- Nature du projet : est-ce une installation physique, une application spécifique à une plateforme ou navigable en ligne ?
- Comment est produit le rendu sonore ? Par le biais de synthèse sonore ou d'échantillonnage ?
- Quelle est la frontière entre la composition et la performance musicales déterminée par le système ? Le rendu sonore est-il en temps réel ou différé ?

- Quelles modalités d'interaction sont proposées ? Des capteurs portés ou manipulés sont-ils nécessaires ? Ou une interaction moins invasive et à distance est-elle choisie ?
- Combien d'utilisateurs peuvent interagir simultanément avec le système ? Un seul ? Plusieurs ?
- Y-a-t-il plusieurs types d'utilisateurs impliqués à différents niveaux d'interaction avec ces systèmes : les utilisateurs directs, le public, des DJs ou "maîtres de cérémonie" (MC's)... ?

2.3. Quelques exemples de systèmes multi-utilisateurs de composition musicale interactive

Suivant la taxonomie dégagée juste avant, nous allons caractériser particulièrement les travaux proposant des installations (si possible multi-utilisateurs) qui permettent de produire une composition sonore (si possible musicale) assistée et basée sur échantillons sonores.

Dance Jockey permet la manipulation en temps réel de musique selon les mouvements corporels de danseurs [4]. Des capteurs portés sont requis pour l'analyse des mouvements, ce qui peut limiter le nombre d'utilisateurs simultanés et par leur nature invasive réduire la spontanéité lors de la phase de découverte de l'installation (les visiteurs devant au préalable "s'habiller" de costumes de capteurs avant de pouvoir interagir).

SoundTorch propose de naviguer dans une carte sonore à l'aide d'une télécommande Nintendo Wii similairement à la découverte progressive d'un espace réel sombre à l'aide d'une lampe torche [9]. Ce projet est initialement une application pour ordinateur personnel, mais pourrait devenir une installation de navigation collective en augmentant le nombre de télécommandes comme prévu pour la console de jeux pour laquelle ces télécommandes ont été conçues.

BeatScape [1] fournit une interface utilisateur plus complexe : un premier groupe de performeurs utilise une interface tangible sur table qui permet de positionner des échantillons sonores qui seront déclenchés par un second groupe de performeurs utilisant des télécommandes Wii, ce qui permet une composition collaborative à divers niveaux d'assignements de tâches.

LoopJam se différencie de ces travaux en proposant une interface utilisateur non invasive grâce à une caméra Kinect et en fournissant une organisation automatique des échantillons sonores par similarité de timbre (et par incidence majoritairement par instrument) suite à l'analyse par traitement de signal de leur contenu.

3. IMPLÉMENTATION

3.1. Architecture

La figure 1 donne une vue d'ensemble de la configuration de l'installation : une carte sonore 2D est projetée sur un mur vertical, plusieurs visiteurs ou joueurs se déplacent sur le sol et leur position correspond à des curseurs de navigation sur la carte affichée.



Figure 1. Photographie de l'installation avec 3 participants.

La figure 2 détaille l'architecture : 4 haut-parleurs sont disposés autour de la zone de captation et fournissent un son spatialisé, les participants sont captés à l'aide d'une caméra Kinect.

3.2. MediaCycle, une suite logicielle pour l'organisation de contenu multimedia par similarité

Nous développons depuis trois ans MediaCycle, une suite logicielle pour l'organisation par similarité de contenu multimedia. Les types de média supportés actuellement sont : audio (des échantillons aux pistes musicales [5]), images, video (particulièrement pour une installation pour explorer des vidéos de danseurs [6]) et texte (en tant que métadonnée pour d'autres types de média ou type à part entière). La suite logicielle propose des implémentations compatibles autant pour des ordinateurs isolés que des applications client/serveur, avec un but commun de fournir une représentation multi-dimensionnelle, visuelle et abstraite, d'une "médiathèque". Son architecture est modularisée par type de média ; par algorithmes d'extraction de caractéristiques dédiées, de classification (l'algorithme standard "k-moyennes" est principalement utilisé), et de calcul de positions de noeuds-media dans une représentation réduite à deux dimensions.

LoopJam et une application multi-plateforme créée avec la suite logicielle MediaCycle [5]. Les fichiers sonores représentés par des noeuds visuels peuvent être parcourus rapidement avec un retour sonore instantané.

3.2.1. Auto-organisée par : similarité de timbre, instrument

Pour organiser les audiothèques visuellement selon des caractéristiques haut-niveau comme le timbre, la rythmique et l'harmonie ; nous avons implémenté plusieurs algorithmes de caractéristiques bas-niveau : spectrales (centroïde, entropie, variation, MFCC ou coefficients cepstraux sur l'échelle de Mel, taux de passage au niveau moyen), perceptuels (intensité perçue), temporels (temps d'attaque logarithmique, fréquence et amplitudes de modulation d'énergie).

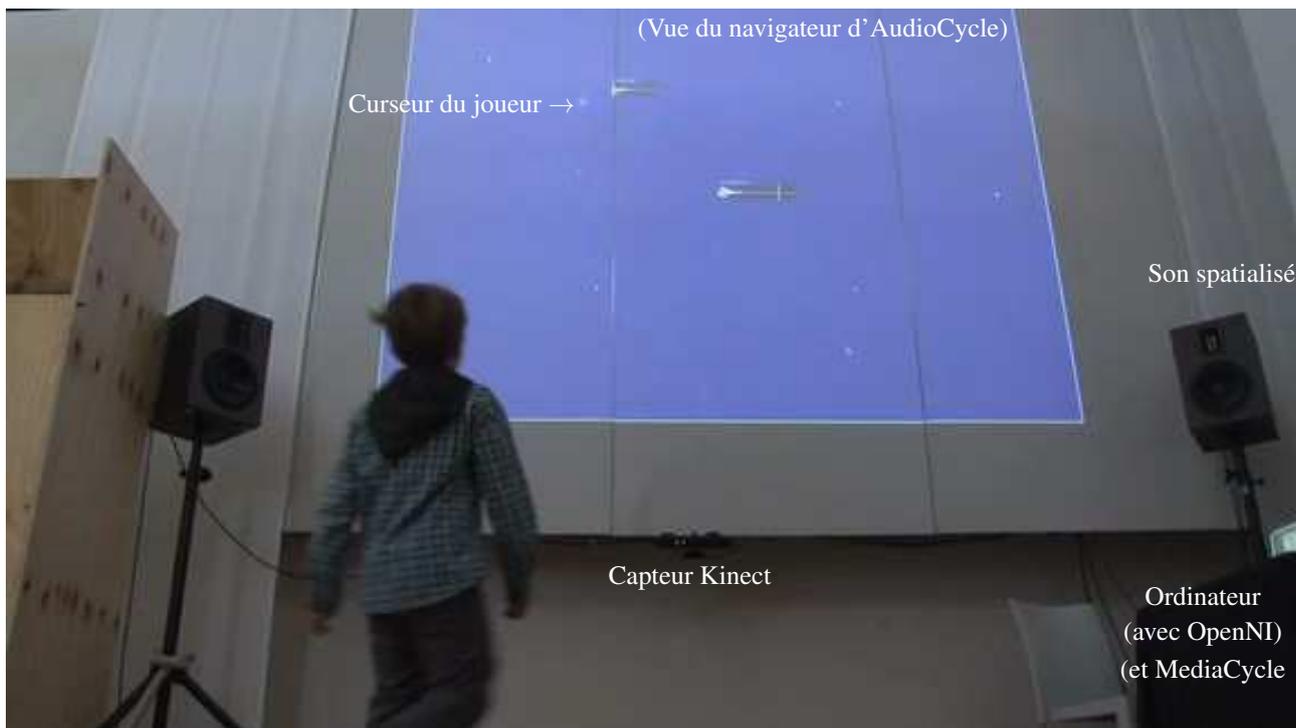


Figure 2. Photographie de l'installation annotée avec des descripteurs de son architecture.

En favorisant un choix de boucles mono-instrumentales, l'analyse du timbre permet donc de classer les sons par "instrument".

3.2.2. Synchronisée sur un tempo commun

Les collections d'échantillons sonores usuellement importées dans l'installation LoopJam utilisent le format ACID, pour lesquelles le tempo est encodé en tant que métadonnée, même si d'autres formats audio sont supportés. Dans le cas du format ACID, LoopJam synchronise automatiquement la vitesse de lecture des échantillons à un tempo commun à l'aide d'un vocodeur de phase.

3.2.3. Spatialisé dans le plan horizontal

Le moteur audio créé pour MediaCycle utilise la librairie OpenAL pour le rendu sonore spatialisé. Les sons sont spatialisés dans le plan horizontal, par défaut la configuration standard est une simple stéréophonie, mais compatible aussi avec une configuration quadriphonique utilisée pour LoopJam pour une meilleure immersion tout en gardant l'installation transportable.

3.3. Une interaction dite naturelle, par les gestes

Nous avons implémenté un contrôle réseau de la suite logicielle MediaCycle par le protocole OpenSoundControl (OSC) initialement pour utiliser des contrôleurs standard pour la navigation inter- et intra-media : molettes rotatives, souris 3D, joysticks à retour d'effort... [7]. Nous avons mis à jour cet interfaçage pour supporter plusieurs

curseurs de navigation et ainsi permettre à plusieurs utilisateurs d'interagir simultanément avec la carte sonore. Nous avons utilisé une caméra Kinect qui capte la profondeur couplé à la librairie OpenNI pour segmenter la scène visuelle 3D d'utilisateurs.

Une option de l'installation LoopJam est de modifier le tempo des échantillons sonores par l'analyse des mouvements des utilisateurs. Nous avons implémenté un algorithme qui mesure la rythmique que les utilisateurs battent. La vitesse de la main gauche est extraite de la représentation du squelette des participants. Nous calculons la somme de ces vitesses et extrayons le maximum de puissance de l'analyse spectrale du signal sur une fenêtre de 2 secondes. La fréquence obtenue est moyennée par un filtre passe bas, cette valeur obtenue contrôlant le vocodeur de phase implémenté dans MediaCycle.

L'utilisateur faisant les mouvements les plus amples influence le tempo majoritairement.

3.4. Collections de boucles audio utilisées dans LoopJam

Parmi les collections de boucles audio ayant été utilisées dans LoopJam :

- des librairies de boucles au format ACID, commerciales mais libres de droit pour la retransmission, par Zero-G (par exemple le DVD *Pro Pack*) et Sony Creative Software (par exemple *8-bit Weapon : A Chiptune Odyssey*) ;
- des collections libres telles que la librairie de boucles libres One Laptop Per Child (OLPC).

4. DISCUSSION

Cette première version de notre installation a été exposée à Seneffe en Belgique pour l'évènement Arts & Sciences les 21 et 22 mai 2011. Elle a également été installée à Paris au Centre Wallonie-Bruxelles, du 22 septembre au 23 octobre 2011. LoopJam a été sélectionné pour le concours artistique du Network & Electronic Media (NEM) Summit à Turin, Italie, les 25-27 septembre 2011, et classé par le jury dans le top 5.

Nous discutons ci-après des choix initiaux après un examen et interrogation des visiteurs lors de la première exposition.

4.1. L'organisation par similarité aide-t-elle ?

L'organisation par similarité basée sur le contenu est essentielle pour cette installation pour deux raisons. Premièrement, comme nous avons choisi de classer les sons par similarité de timbre, le chemin parcouru par chaque joueur dans l'installation et par incidence dans la représentation des sons est consistante ; les participants s'attendent à trouver des sons du même instrument dans une même zone, à la manière de la localisation des instrumentistes dans un orchestre classés par type d'instrument (sections de cordes, de cuivres, etc...). Deuxièmement, l'organisation automatique facilite la tâche de préparer le contenu de chaque collection de sons parcourue dans LoopJam, en opposition à un classement totalement manuel.

4.2. Interaction directe ou indirecte ?

Nous avons initialement prévu de projeter la carte sonore sur le sol pour permettre une interaction directe [3], les participants auraient alors vu la représentation visuelle à leurs pieds, en correspondance spatiale directe avec la représentation de la zone d'interaction. Cependant ceci aurait impliqué un certain nombre d'inconvénients en comparaison avec la configuration actuelle : le fait que les gens orientent leur regard au sol incite moins au contact social, des noeuds visuels pourraient être masqués par occlusion avec des participants, une configuration robuste vis-à-vis de l'occlusion aurait requis du matériel bien moins transportable (par exemple une plateforme horizontale avec rétro-projection).

4.3. Quelle est l'expressivité des visiteurs, préparateurs et responsables de l'installation ?

A travers cette installation, nous redéfinissons la relation entre les visiteurs et les préparateurs de l'installation. L'"audience" peut participer à la performance musicale en dansant, son comportement et ses mouvements affectant le choix des échantillons sonores ainsi que les paramètres de lecture (tempo, volume). Le préparateur de l'installation peut insuffler sa sensibilité artistique à travers le choix des collections sonores présentées [10], d'autant plus grande mesure quand les sons sont créés par le préparateur-même. Les DJs faisant attention aux signes de

désengagement du public [8], le responsable de l'installation peut également influencer sur l'ambiance générale pendant la performance en variant les paramètres de l'organisation des collections : choix de collections, navigation plus précise dans une collection par zoom ou rotation spatiale. S'il s'agit de la même personne, le préparateur et responsable de l'installation peut donc assurer une certaine qualité de la performance musicale, avant et pendant, laissant s'exprimer les participants en leur laissant une variabilité contrôlée et assistée des paramètres à leur disposition.

4.4. Comment maintenir la satisfaction des visiteurs ?

Notre méthode actuelle pour garder les participants satisfaits et apprécier l'installation est majoritairement de varier fréquemment la collection musicale exposée pendant la durée de l'installation, ce qui est automatisé avec un séquençement. Nous avons choisi de ne pas ré-organiser une collection pendant une performance car nous pensons que les utilisateurs se constituent une carte mentale des sons similaire à leur représentation visuelle. Comme les visiteurs consacrent un temps assez réduit à l'installation, ils pourraient être découragés et perturbés par les changements de représentations, à la manière d'un instrumentiste auquel on imposerait un accordage non-standard en performance. Les méthodes de navigation telles que le zoom et la rotation permettent un juste milieu : affiner la vue tout en conservant le même modèle. Les rotations permettent également aux visiteurs d'intervir les instruments.

5. CONCLUSIONS, PERSPECTIVES

Nous avons proposé une installation qui mêle informatique musicale et interaction multi-utilisateurs pour composer et "performer" de la musique en s'amusant. L'auto-organisation par similarité de timbre basée sur le contenu accélère résolument l'import de collections de boucles sonores pour la préparation ou personnalisation de l'installation : et fournit une représentation spatiale des boucles sonores organisées par instrument à la manière des instrumentistes d'un orchestre.

Testée au cours de trois expositions, cette installation pourrait être transposée à d'autres types de lieux comme les boîtes de nuit, les institutions socio-culturelles, des parcs d'attraction...

Le nombre de boucles dans chaque collection, l'espace résultant entre chaque noeud représentant une boucle, et la taille de ces noeuds pourraient rendre l'installation trop sensible aux déplacements de faible amplitude des visiteurs. Des possibilités pour compenser cet inconvénient seraient d'adapter la taille des collections et la dimension de l'espace sensible en utilisant plusieurs caméras combinées et fusionnées.

Si nous pouvons déjà extraire des représentations du squelette de chaque participant dans la scène 3D grâce à la caméra, nous pourrions envisager de parfaire la détection de rythmique individuellement afin de permettre le

changement de tempo par boucle.

Nous pensons à d'autres façons de personnaliser l'installation. Les téléphones mobiles des participants pourraient être utilisés pour synchroniser avec leurs profils sur des sites de recommandation tels que LastFM et récupérer l'historique de leurs goûts musicaux, alors que les téléphones de l'audience pourraient servir à voter sur le choix des collections de sons, ou proposer aux joueurs de "jammer" ensemble, voire de céder leur place ; éventuellement par manifestation vocale [2]. Nous pourrions remplacer les curseurs visuels de chaque participant par des avatars générés grâce à la caméra.

Enfin, à l'aide d'analyse informatisée des mécanismes d'attention sociale, nous pourrions améliorer la jouabilité de l'installation : donner plus de poids aux participants aux comportements saillants, inviter d'autres gens parmi l'audience à participer.

6. REMERCIEMENTS

numediart est un programme de recherche autour des technologies des arts numériques, financé par la Région Wallonne, Belgique (grant N°716631).

Nous voudrions remercier notre ancien collègue Damien Tardieu pour avoir suggéré le concept de cette installation.

Notre travail aurait moins de visibilité sans les démos vidéo professionnelles de Laura Colmenares Guerra.

7. REFERENCES

- [1] A. Albin, S. Senturk, A. V. Troyer, B. Blosser, O. Jan, and G. Weinberg. Beatscape and a mixed virtual-physical environment for musical ensembles. In A. R. Jensenius, A. Tveit, R. I. Godøy, and D. Overholt, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 112–115, Oslo, Norway, 2011.
- [2] L. Barkhuus and T. Jørgensen. Engaging the crowd : studies of audience-performer interaction. In *CHI '08 extended abstracts on Human factors in computing systems*, CHI EA '08, pages 2925–2930, New York, NY, USA, 2008. ACM.
- [3] S. K. Card, J. D. Mackinlay, and G. G. Robertson. The design space of input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'90)*, pages 117–124, 1990.
- [4] Y. de Quay, S. A. v. D. Skogstad, and A. R. Jensenius. Dance jockey : Performing electronic music by dancing. *Leonardo Music Journal*, 21 :11–12, 2011.
- [5] S. Dupont, C. Frisson, X. Siebert, and D. Tardieu. Browsing sound and music libraries by similarity. In *128th Audio Engineering Society (AES) Convention*, London, UK, May 22-25 2010.
- [6] C. Frisson, S. Dupont, X. Siebert, and T. Dutoit. Similarity in media content : digital art perspectives. In *Proceedings of the 17th International Symposium on Electronic Art (ISEA 2011)*, Istanbul, Turkey, September 14-21 2011.
- [7] C. Frisson, S. Dupont, X. Siebert, D. Tardieu, T. Dutoit, and B. Macq. DeviceCycle : rapid and reusable prototyping of gestural interfaces, applied to audio browsing by similarity. In *Proceedings of the New Interfaces for Musical Expression++ (NIME++)*, Sydney, Australia, June 15-18 2010.
- [8] C. Gates, S. Subramanian, and C. Gutwin. Djs' perspectives on interaction and awareness in night-clubs. In *Proceedings of the 6th conference on Designing Interactive systems*, DIS '06, pages 70–79, New York, NY, USA, 2006. ACM.
- [9] S. Heise, M. Hlatky, and J. Loviscach. Soundtorch : Quick browsing in large audio collections. In *125th Audio Engineering Society Convention*, number 7544, 2008.
- [10] T. Rodgers. On the process and aesthetics of sampling in electronic music production. *Organised Sound*, 8(3) :313–320, December 2003.