LoopJam: turning the dance floor into a collaborative instrumental map

Christian Frisson, Stéphane Dupont, Julien Leroy, Alexis Moinet,
Thierry Ravet, Xavier Siebert, Thierry Dutoit
University of Mons (UMONS), TCTS lab
Boulevard Dolez 31 B-7000 Mons, Belgium
name.surname@umons.ac.be

ABSTRACT

This paper presents the LoopJam installation which allows participants to interact with a sound map using a 3D computer vision tracking system. The sound map results from similarity-based clustering of sounds. The playback of these sounds is controlled by the positions or gestures of participants tracked with a Kinect depth-sensing camera. The beat-inclined bodily movements of participants in the installation are mapped to the tempo of played sounds, while the playback speed is synchronized by default among all sounds. We presented and tested an early version of the installation to three exhibitions in Belgium, Italy and France. The reactions among participants ranged between curiosity and amusement.

Keywords

Interactive music systems and retrieval, user interaction and interfaces, audio similarity, depth sensors

1. INTRODUCTION

The purpose of the LoopJam installation is to gather people to collaboratively build a live musical atmosphere. A two-dimensional "sound map" is created by our software which analyzes sounds, as well as musical loops, and groups them by similarity. Wandering through the installation, each participant explores this sound map and activates sounds by simple gestures matching the desired tempo. The playback of each sound is synchronized by our software, allowing a collaborative audio composition created by all participants.

In section 2 we provide the context of this installation and quickly describe related works. In section 3 we detail our architecture and implementation choices. In section 4 we discuss the preliminary conclusions we produced after watching people use the installation. After concluding in section 5, we propose upcoming possible ways of improving this installation in section 6.

2. RELATED WORKS

2.1 Domains of application

The core objective of the LoopJam installation is to allow users to interactively create musical content.

The earliest computerized systems that can be considered as its first ancestors, since the 1950's, are sound synthesis environments, giving access to parameters of audio signal processing algorithms, and multitrack audio sequencers, directly inspired from the tangible techniques of cutting and pasting magnetic tapes. These systems proposed user interfaces of variable "friendliness", initially with punched cards or command line user interfaces, in delayed time due to computer processing limitations; later with graphical user interfaces paired with mice, keyboards and optional controllers fitted with potentiometers, sliders, and so on... with a more seamless interactivity, in real-time. A more recent interest moves towards tangible user interfaces and multitouch devices, to regain the lost sense of embodied interaction with music, but WIMP interfaces are still mainly used in professional studios for music composition.

The advent of the personal computers shifted the usage of these systems not exclusively from experts and researchers, but as well to consumers and amateurs. Crossovers between music performance and the video game industry, particularly with the Guitar Hero or Rock Band franchises, extended the interest for music creation towards more users, not necessarily skilled as musicians, expecting to blend creativity with simplicity.

2.2 A taxonomy

LoopJam combines a range of widespread or emerging technologies, including audio and music signal processing, immersive audio, music information retrieval, audio browsing, 3D computer vision, gestural control, behavioral recognition.

Related works can be classified regarding the following characteristics:

- Nature of the project: is it a physical installation, a desktop application, an online application?
- How is the sound rendering produced? By means of sound synthesis? Or sound sampling?
- How is boundary between music composition and performance delimited by the system? Is the sound rendered in real or delayed time considering the user interaction?
- Which interaction modalities are proposed? Are wearable sensors required? Or is a less invasive free-form interaction chosen?
- How many users can interact with the system? A single one? Many?
- Are there multiple types of users involved with different levels of interaction with the systems: direct players, audience, disk jockeys or "masters of ceremony" (MC's)...?

NIME'12, May 21 – 23, 2012, University of Michigan, Ann Arbor.

2.3 Some examples of multi-user interactive audio composition systems

Using the aforementioned taxonomy, we will characterize mostly works consisting in sample-based installations where multiple participants can interact with the system to produce a computer-aided musical composition.

Dance Jockey allows to manipulate music in real-time depending on dancers' bodily movements [4]. Wearable sensors are required for the analysis of the movements, which might limit the number of participants and might invasively decrease the spontaneousness during the discovery phase of the installation (participants first requiring to dress up with the motion capture suit before being able to interact with the installation).

SoundTorch proposes to browse a sound map using a Nintendo Wii remote controller similarly to discovering progressively a real 2D space with a light torch [9]. While it is by nature a desktop application, increasing the number of remote controllers, what is feasible with the game console to which these controllers were initially dedicated, might turn this project into an installation where players would browse the sound map collaboratively.

BeatScape [1] proposes a more complex user interface where a first set of performers uses a tangible tabletop interface to place sound samples which are triggered by a second set of performers using Wii remotes; what allows a collaborative music composition with different levels of task assignments.

LoopJam differentiates itself from these works by proposing a non-invasive user interface based on a Kinect camera and by providing content-based auto-organization of sounds by similarity of timbre (sorting by instrument).

3. IMPLEMENTATION

3.1 Architecture

Figure 1 gives a general overview of the installation setup: a 2D sound map is projected on a vertical wall, several participants or players move on the floor and their position is mapped to browsing cursors on the screened map. Figure 2 details the architecture: four speakers are located around the sensing zone to provide surround sound, participants are tracked using a depth sensor.

Figure 1: Picture of the installation with 3 participants.

3.2 MediaCycle, a framework for multimedia content organization by similarity

We have been developing since three years MediaCycle, a framework for the organization by similarity of multimedia content. The currently supported media types are: audio (from loops to music tracks [5]), images, video (notably in an installation for browsing dancers videos [6]) and text (as metadata to other media types or by itself). The framework proposes implementations of both desktop and client/server applications, with a shared goal of providing a multi-dimensional representation of a media library, visual or abstract. Its architecture is modularized along media types and related feature extraction algorithms, clustering algorithms (the standard K-Means algorithm is mainly used) and computation of media node positions in a representation scaled to 2D.

LoopJam is a desktop application created with the MediaCycle framework dedicated to audio content [5]. Audio files represented by visual nodes can be accessed quickly and browsed by "hovering" with an instant sound feedback.

3.2.1 Auto-organized by: timbral similarity, instrument

To organize audio libraries visually along high level features such as timbre, rhythm and harmony; we implemented many standard lower-level audio feature algorithms: spectral (centroid, spread, variation, flatness, decrease, MFCC, zero crossing rate), perceptual (loudness), temporal (log attack time, energy modulation frequency / amplitude).

3.2.2 Synchronized to a common tempo

The loop databases usually loaded in the LoopJam installation are using the ACID format, where the tempo is encoded as metadata for optimal reliability, although standard audio format are supported as well. In the case of ACID databases, LoopJam automatically synchronizes the playback of loops and adapts the playback speed of each to a common tempo using a phase vocoder.

3.2.3 Spatialized in the user plane

The audio engine created for LoopJam uses the OpenAL library for sound playback and spatialization. The sounds are spatialized in the user plane, the default setting being for a stereo setup, but is compatible as well with a quadriphonic setup used for LoopJam for an enhanced immersion while keeping the installation transportable.

3.3 Natural interaction with gestures

We implemented networked control bindings in the MediaCycle framework with OpenSoundControl (OSC) initially to make use of off-the-shelf devices for the inter- and intra-media browsing (jog wheels or rotary controllers, 3D mice or force-feedback joysticks...) [7]. We updated these bindings and added multiple pointer support so as allow multiple users to interact with the sound map. We used a Microsoft Kinect depth sensing camera and the OpenNI library to segment users in a 3D scene.

An option of the LoopJam system is to modify the tempo of the played audio loops by analyzing the movement of the users. We implemented a function that measures the rhythm that the users beat out. The speed of the left hand is extracted from the skeleton representation of the participants. We sum these speeds and we extract the frequency by computing the maximum power in the spectral analysis of this signal over 2-second windows.

The obtained frequency is smoothed by a lowpass filter; this value is used to control the tempo of the audio engine output with the phase vocoder algorithm implemented in MediaCycle.

The user who makes the larger movements influences the music tempo the most.

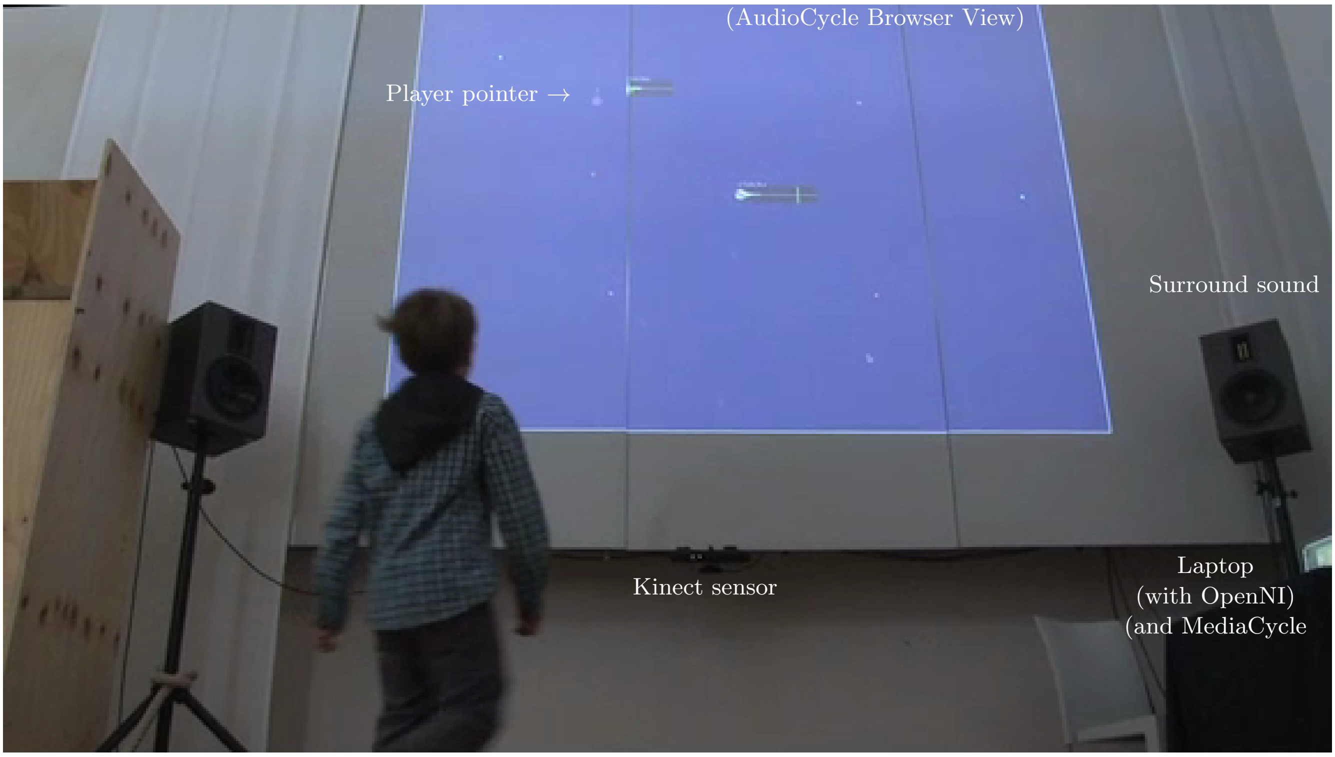

Figure 2: Picture of the installation, annotated to describe its architecture.

3.4 Audio databases used with LoopJam

The loop databases that have been used in LoopJam include:

- commercial but royalty-free ACID loop libraries from Zero-G (for instance the *Pro Pack* double DVD) and Sony Creative Software (for instance *8-bit Weapon: A Chiptune Odyssey*);
- opensource collections such as the One Laptop Per Child (OLPC) Free samples library.

4. DISCUSSION

This first prototype of our installation has been shown in the Arts & Sciences exhibition in Seneffe, Belgium, May 21st and 22d, 2011. It was also featured at an exhibition in Paris, at the Centre Wallonie-Bruxelles, from September 22nd to October 23rd, 2011. LoopJam was selected at the Art Contest of the 2011 Network & Electronic Media (NEM) Summit in Torino, Italy, September 27-29, where it was ranked by the jury within the top 5 of all proposals.

We discuss initial choices after having examined and interviewed participants during the first exhibition.

4.1 Does organization by similarity help?

Organization by content-based similarity is essential for this installation due to two main reasons. First, since we chose to sort sounds by timbral similarity, it provides consistency in the path browsed by each player and the resulting sound rendering, expecting to find sounds of the same instrument or instrument family nearby, similarly to the fact that instrumentalists in a band or orchestra are grouped by instrument family and play a dedicated instrument. Additionally, the automated organization eases the tasks of choosing the contents of each sound library, versus manual loop sorting.

4.2 Direct or indirect interaction?

We initially planned to project the sound map on the floor to allow a direct interaction [3], as participants would see the visual representation of the loops played back located under their feet. However it would have brought several drawbacks in comparison to the current setup: people looking at their feet usually reflects timidity, other participants may occlude each others nodes visually, an occlusion-proof setup would have required the equivalent of a large rear-projected multitouch table quite less transportable.

4.3 How do players and installation curators express themselves?

With this installation, we refine the participants/artist relationship. The audience participates to the music performance by dancing, its behavior and movements affecting the choice of musical loops and hence the resulting mix, as well as the overall playback parameters (tempo, amplitude...). The DJ or curator responsible for the installation introduces her/his artistic sensibility by the choice of sound libraries [10], even more when sounds created by the curator herself/himself are featured, and can affect moods during the performance by altering meta parameters affecting the navigation/browsing: changing sub-libraries, re-organizing the library display spatially (zoom, rotations...) and in terms of contents (groups of sounds assorted by different facets such as timbre, rhythm...). Additionally, DJs watch for signs of disengagement in the participants [8]. The curator may therefore ensure a certain quality in the global performance (during the performance, and before, while choosing sound libraries), while the audience participates to the playback, whereas seamless sound synchronization and minimal automated musicianship is taken care of by the installation audio rendering back-end.

4.4 How to keep participants satisfied?

Our initial method for keeping participants to enjoy the installation was to change the sound data regularly, yet at repeated time intervals and using different libraries sequenced in no particular order. We chose not to remap or recluster a library during performances since we believe participants construct a mental model of each library by exploring the map as provided at its first state and would become annoyed to start from scratch again with the same content, however rotations of the visual map could provide an intermediary solution.

5. CONCLUSION

We proposed an installation that blends music information retrieval and multi-user interaction to compose and perform music in an amusing way. Auto-organization by content-based similarity of timbre definitely speeds up the integration of sample libraries for the preparation or customization of the installation, and provides a spatial layout of loops sorted by instrument similarly to the position on stage of players of traditional music bands.

6. PERSPECTIVES

Tested during three exhibitions, this installation could be sized to other venues such as nightclubs, socio-cultural institutions, amusement parcs...

The number of samples in the library and the projected sizes of nodes and spacing between them might make the installation too sensitive to small movements of participants. One possibility to overcome this problem is to provide a wider sensing zone by the fusion of information from multiple depth sensors.

We can already map virtual skeletons to segmented people in the 3D scene using the depth sensor, however we didn't make use of this feature so far. By detecting the frequency of beating of each/all participant(s), a supplementary control of the tempo of each/all played back sounds could be considered.

We foresee several ways for personalizing the installation to participants. Mobile phones of participants could be used to synchronize their potential profiles of music recommendation systems such as LastFM so as to retrieve the history of their musical tastes, while the non-participating audience could use the same devices (or a more natural means such as shouting or other vocal manifestations [2]) to vote for a change in participants or sound libraries. We plan to replace visual pointers of each participant with avatars generated from the depth-sensing camera.

Using latest advances in social attention analysis, we could introduce a game feel to the installation: give more weight to salient participants, invite people from the audience to participate.

7. ACKNOWLEDGMENTS

`numediart` is a long-term research program centered on Digital Media Arts, funded by Région Wallonne, Belgium (grant N°716631).

We would like to thank our past colleague Damien Tardieu for pitching the concept of this installation.

Our work would be less visible without Laura Colmenares Guerra's professional demo videos.

8. 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 nightclubs. 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.