YARMI: an Augmented Reality Musical Instrument

Tomás Laurenzo **Ernesto Rodríguez** **Juan Fabrizio Castro**

Universidad de la República
Herrera y Reissig 565, 11300
Montevideo, Uruguay.

laurenzo, erodrig, jfcastro@fing.edu.uy

Abstract

In this paper, we present *YARMI*, a collaborative, networked, tangible, musical instrument, being developed at the Universidad de la República in Uruguay. *YARMI* operates on augmented-reality space (shared between the performers and the public), presenting a multiple tabletop interface where several musical sequencers and real–time effects machines can be operated. We first introduce the instrument's concepts and design, and then some implementation details and the performance setup.

Keywords: Interactive music instruments, audio visualization, visual interfaces, visual feedback, tangible interfaces, computer vision, augmented reality, collaborative music, networked musical instruments, real-time musical systems, musical sequencer, step sequencer, continuous sequencer.

1. Introduction

The field of computer based musical instruments, while very active and vital, has produced several instruments that somehow lack playability or expressiveness in traditional musical terms.

While the creation (or, better, the composition [1]) of new musical instruments can provide composers and performers with tools for new music and musical languages, the instruments produced are often too simplistic (like the rather naïve new incarnations of traditional step sequencers and drum machines), lack of playability (often due to the delegation of too many performative decisions to the instrument, not providing an effective fly-by-wire alternative), or are too different from traditional instruments (i.e. those with a social agreement on how they should sound, how they should be played or how they relate to others instruments output).

In the words of MIT's Tod Machover, "we risk having the whole field of interactive expression become an historical curiosity, a bizarre parallel to the true pulse of cultural growth".

While the eclecticism of new musical instrument production is easy to note, also "the technological convergence is diminishing the identity of the music technology" [12]: in an ever-more technologically imbued world, the mere fact of using new technologies does not turn an instrument into something interesting.

1.1 Our design goals

YARMI's idea (and its name) emerged from a discussion in our lab: which are the aspects that make an instrument truly playable and not Yet Another Ridiculous Musical Interface? What design patterns can be applied to achieve playability, engagement and musical sense? And also, if there is no social knowledge on the instruments' use, how can the public relate to the performance[1]?

In order to build an instrument that meets those expectations we decided to use two design patterns: (directly) mapping sound to physical objects, and traditional music sequencers.

A rough division of computer music control, based on the immediacy of the sonic response to the interaction, can divide controllers in sequencers and continuous –gestural– controllers.

Music sequencers provide a means to describe a sequence of sounds that are to be produced by the system, and play a central role in computer music creation [2].

Continuous controllers, on the other hand, provide a direct control of the sound being produced, allowing the performer to trigger sounds or modify some of its qualities, more in the vein of a traditional musical instrument.

Both design patterns are extremely important. While sequencers are the traditional tool construct digital music, direct manipulation approaches potentiate users' engagement and (real-time) expressiveness.

In addition, both sequencers and direct manipulation gestures can offer a very gentle learning curve to the performers-to-be while being able to be easily interpreted (that is, to establish a correspondence between the gestures and the produced music) by the audience during a performance.

In effect, musical sequencers are standards that do not pose a metaphor but constitute a recognizable (and comprehensible) interface themselves.

In the same vein, physically based, direct manipulation interaction, constitute an established paradigm in tangible user interfaces, with the successful examples of the

[1] A major concern of us is how the public can decode the performer gesture's relating them to the sonic output.

Reactable [6] (a very strong source of inspiration for us[2]), *ISS Cube* [11], *Audiopad* [10], or *Block Jam* [8].

This interaction style allows users to *feel* that they are operating directly with the objects presented to them [9], also allowing for an easy deconstruction of the performance from the audience.

2. The instrument

YARMI is a collaborative musical and –to a lesser extent– visual instrument being developed at the CECAL research group[3] of the Instituto de Computación of the Universidad de la República[4] of Uruguay.

It was designed to offer tangible, direct, multi-modal, and multi-user interaction, with a shared (between the performers and the public) performance space with explicit visual representation.

2.1 Stations

YARMI is a multi-user, distributed instrument; or rather, an ensemble of synchronized instruments, operating under client-server architecture.

Each performer operates a *station* and *YARMI* is comprised of an arbitrary number of *stations* and one server.

A *station* consists of a table (or any flat surface) with a zenithal camera mounted[5] on and a visual projection showing the *screen-space*, an augmented version of the station's table.

On each *station*'s table, users[6] can put tangibles – wooden tokens with fiducial markers–, which are recognized by the *station* and provide the only mean of interaction.

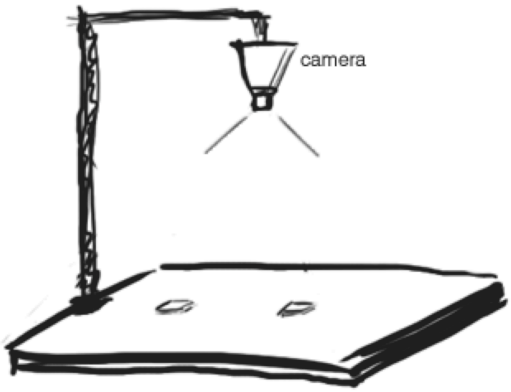

Figure 2. A station's table with the zenithal camera mounted.

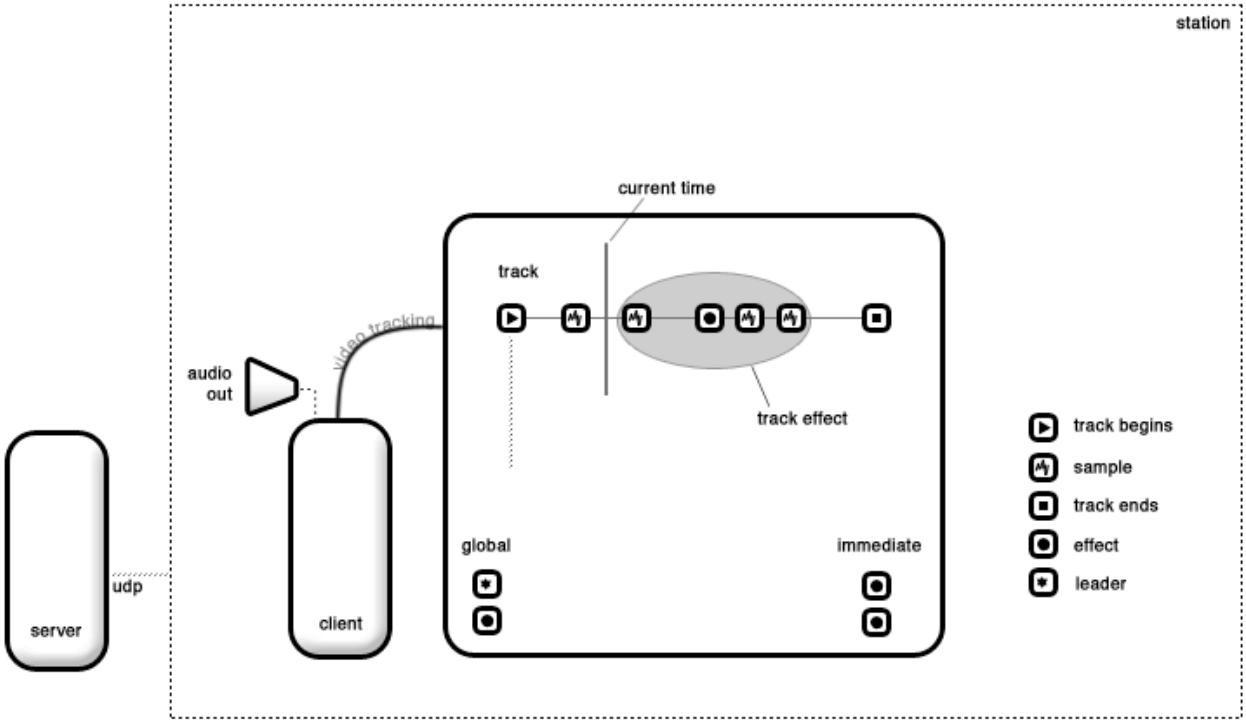

Figure 1. Schematics of one *station*, with some aspects of the *screen space* superimposed, and *YARMI's* server.

2 Another major source of inspiration, albeit perhaps more indirect, was Julian Oliver's levelHead and its augmented reality based interaction.

3 http://www.fing.edu.uy/inco/grupos/cecal

4 http://www.universidad.edu.uy

5 We have designed a camera mount, so any standard table can be used for a station. We hope this will come handy when performing at public spaces.

2.2 Setup

Each station has its own visual representation showing the augmented surface, which we call *screen space*, but the table remains with no augmentation at all: for both the performer and the audience, it is just a flat surface with some wood pieces on it.

The locus of attention of both the performers and the public is the *screen space*. This real-time projection of the augmented surface shall be set up so that the audience stands between the performers and the images, visible by everyone and providing an explicit representation of the performers gesture and the different *stations'* visual feedback.

2.3 Interaction

Each station is divided into three different zones named, *track zone*, *effects zone* and *global zone*, which we will now describe.

The *track zone* is an implementation of a multi-track sequencer, where users can create tracks and add samples and local effects to them.

To create a new track, the performer must add two tokens, one marking its beginning and one marking its end (see Figure 1).

Once a track is created, users can add new tokens to it indicating samples[7] to be played, or local effects to be applied.

In every case, the rotation of the token controls the main parameter of the effect or the sample's pitch.

The *effects zone* (which we also refer to as the *immediate* zone) presents the likes of a sound effects machine. Each token put on it triggers an immediate response.

If the performer adds a token representing a sample, the station starts to play it, looping it as long as the token is present (token rotation always controls the sample's pitch).

If a token representing an effect is added, the effect is applied immediately to the *station*'s output, i.e. the mix of all its tracks and local effects.

If many effects are added, they are applied respecting the order in which they were added to the zone.

Finally, the *global zone* is the *settings* zone, where users can add tokens that modify the *station* or the ensemble behavior.

2.4 In-*station* synchronization

Being a multi-track and multi-user instrument, synchronization between tracks and between stations is fundamental to produce coherent music.

Each track is automatically synchronized so they all start playing at the same time, but, as they can have different lengths, the first track that is created in the leader station (see next subsection) –called the main track– defines the main length (with its speed depending on what is set on the *global* zone).

The *station* always assumes that the main track is 32[8] beats length.

If the performer creates a very short or very long track, for example one of approximately one quarter of the main track length, this is detected and then the track is played four times per bar (the recognizable lengths are half, one quarter and twice the length of the main track).

2.5 *Leader station* and inter-*station* synchronization

As the synchronization between the different *stations* is as important as the synchronization between tracks, we defined that one station is always acting as the *leader station*, and defines when the tracks begin to be played, the performance speed (in BPM), the global volume, etc.

Any station can be the *leader station*. We use a token (the leader token), that, when added to the *global* zone of a *station*, sets it as the leader (the first leader is always the first *station* to connect to the server).

The *leader station* sends its commands to the server, which, in turn, broadcasts them to all the registered *stations*.

2.6 Settings

Several configuration parameters can be modified in performance-time. Each setting has an assigned token and its parameter is controlled by rotating it.

Implemented settings are:

- Velocity (byte): a multiplier of the volume set by the leader.
- Quantization (Boolean): specifies whether the sample's positions in a track are snapped to the closest beat.
- BPM (byte): sets the global bits-per-minute.

Any of these settings affect only the *station* where they are modified, except in the case of the *leader station* where they affect the whole ensemble.

2.7 Visual feedback

Although *YARMI*'s visual feedback is as important as the sound produced, being a project in development, its visual capabilities are in their infancy.

In the current stage the *screen space* shows the table, with the following additions:

- One line for each track.
- One cursor per track showing the current time.
- Several numbers showing the elapsed time, bar count, current main volume and station volume.

6 Depending on the local setup it could be chosen to have more than one performer per station. Our design, however, is focused in the one-performer-per-station setup.

7 As of now, YARMI does not provide any tool for assigning a sample to a token, so this assignment is done in configuration time, before actually using the instrument.

8 This is an off-line configuration parameter.

3. Implementation

YARMI's software was coded in C++ using OpenFrameworks [4] as a programming framework. Several libraries are used, most notably ARToolkitPlus [5] for the recognition and tracking of the tokens' fiducial markers and FMOD [3] for audio playback.

3.1 Software architecture

As we stated before, *YARMI* follows a client-server architecture, with every *station* identical to each other.

Stations have a GlobalState object that models all the *station*'s information. This object is updated by an ARProcessor object (running on its own thread) and is polled by the sound and video managers (see Figure 3).

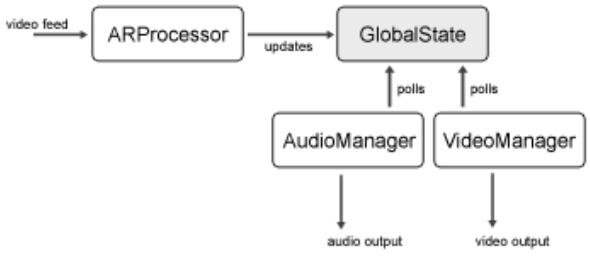

Figure 3. *Station* architecture.

Each time a setting is changed in a *station*, it notifies the server which –if the station is the current *leader*– broadcasts the new settings to all the registered *stations* (if it is not the leader the server ignores the new setting[9].

4. Conclusions

Although *YARMI's* design is in a mature phase, and we have a working prototype, it is a project in development for which much work is yet to be done.

A major milestone yet to be reached is to test *YARMI* in a real performance setup, so far we have used it in our lab, in a controlled environment.

We believe that *YARMI* has some characteristics that can turn it into a capable, and interesting, musical instrument

Its networking approach, with many identical components that synchronize themselves automatically, allow for a confident use (delegating some of the cognitive work of the performance onto the system), while maintaining the performers engagement, independence and expressivity, which, in turn are levered with the inclusion of the *immediate zone*.

This combination of the *track* and *immediate* zones offer the "safety" of computer-sequenced music with the expressiveness of traditional musical instruments.

Finally, we believe that the explicit representation of the instruments' feedback, together with the performance happening on a virtual space external to both the audience and the performers allow the public to decode some of the performance aspects re-installing the lost dimension of virtuosity into the performance.

Virtuosity has traditionally played an important role in live musical performances with an added aesthetic dimension of its own. But, for virtuosity to play that role, the performance details must be understood by the audience.

With *YARMI*, once again, the audience can enjoy not only the sonic output of the instruments but also *how* the sounds are created.

4.1 Future work

Besides improving the instrument's output (specially its visual output), some paths appear worth to follow.

Specifically, we would like to perform additional research in the following directions:

- Active inclusion of the audience: if the public can decode the performers' gestures, the next step is to allow them to actively participate in the performance.
- Geographically distributed performances: having an inherently networked instrument would allow us to explore the relevance of proximity and simultaneity in *YARMI*'s performances.
- More multimodal interaction: we would like also to investigate whether new interaction styles and techniques can be added.

References

[1] Bahn, C and Trueman, D. "Interface, Electronic Chamber Ensemble" Proceedings of the CHI 01 Workshop New Interfaces for Musical Expression (NIME01), Seattle, USA, 2001.

[2] Duignan, M., Biddle, R., & Noble, J. (2005). "A taxonomy of sequencer user-interfaces". International Computer Music Conference (ICMC). Barcelona: Inter-Society of Electronic Arts.

[3] http://fmod.org., web site. Last visited, January 2009.

[4] http://openframeworks.cc, web site. Last visited January 2009.

[5] http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php, web site. Last visited, January 2009.

[6] Kaltenbrunner, M. & Jordà, S. & Geiger, G. & Alonso, M. "The reacTable*: A Collaborative Musical Instrument," Proceedings of the Workshop on "Tangible Interaction in Collaborative Environments" (TICE), at the 15th International IEEE Workshops on Enabling Technologies (WETICE 2006). Manchester, U.K.

[7] Livingstone, D., O'Shea, C. (2005) "Tactile Composition Systems for Collaborative Free Sound", Proceedings of the International Computer Music Conference, Barcelona.

[8] Newton-Dunn H., Nakanon H., Gibson J. "Block Jam: A Tangible Interface for Interactive Music" Proceedings of the

[9] Therefore, all the stations act the same, regardless if they are leaders or not.

Conference on New Interfaces for Musical Expression, Montreal, Canada 2003

[9] Norman, D. (1988) “The psychology of everyday things”, USA: Basic Books

[10] Patten J., Recht B., Ishii H."Audiopad: A Tag-based Interface for Musical Performance” Proceedings of the Conference on New Interfaces for Musical Expression, Dublin, Ireland, 2002.

[11] Quarta, M. “ISS Cube”, exhibited at Ars Electronica, Cybersonica, Bafta Interactive Festival 2003 http://www.somethingonline.org/txt/d_iss.htm

[12] Serra, X. “Technological Innovation in the Current Social Context: Who is really in control?”, keynote at NIME 2008, Genova, 2008