Sync'n'Move: social interaction based on music and gesture

Giovanna Varni, Maurizio Mancini, Gualtiero Volpe, and Antonio Camurri

InfoMus Lab, DIST - Università degli Studi di Genova, Italy
giovanna.varni@infomus.dist.unige.it,maurizio.mancini@dist.unige.it,
gualtiero.volpe@unige.it,antonio.camurri@unige.it

Abstract. In future User Centric Media the importance of the social dimension will likely increase. As social networks and Internet games show, the social dimension has a key role for active participation of the users in the overall media chain. In this paper, a first sample application for social active listening to music is presented. Sync'n'Move enables two users to explore a multi-channel pre-recorded music piece as the result of their social interaction. The application has been developed in the framework of the EU-ICT Project SAME (www.sameproject.eu) and has been presented for the first time at the Agora Festival (IRCAM, Paris, June 2009). In that occasion, Sync'n'Move has also been evaluated by both expert and non expert users.

Key words: active music listening, social interaction, synchronization

1 Introduction

In the Future Internet users will play a crucial role both as individuals and in the social dimension. The User Centric Media concept "implies that the user will become an active member of the overall media chain by generating, distributing and experiencing high-quality media content" [1]. The worldwide spreading of social networks and Internet games bears witness of the importance of the social dimension for such an active participation of the users in the overall media chain.

Nevertheless, many existing multimedia interactive systems and Internet applications are still intended for a single user and social interaction is often neglected. Social networks, indeed, are mainly based on sharing of static textual and audiovisual content, whereas realtime interaction between users, immersiveness, and sense of presence are far to be fully reached.

In the framework of the EU-ICT Project SAME (www.sameproject.eu), novel paradigms of active, embodied, and social listening to music in context-aware mobile applications are explored [5], i.e., paradigms enabling both single and groups of users to actively mould and reshape sound and music content, based on movement, gesture, and social interactions.

Sync'n'Move is a first sample application in the direction of social active listening to music. It enables two users to explore a multi-channel pre-recorded music piece as the result of their social interaction. Users interact through the movements they perform by handling a mobile device. A phase synchronization index is extracted from movement. High-level of synchronization enhances music orchestration and rendering, by adding music sections and rendering features. Sync'n'Move has been presented for the first time at the Agora Festival (IRCAM, Paris, June 2009), as one of the sample applications proposed by the SAME Project. In that occasion, Sync'n'Move has also been evaluated by both expert and non expert users.

The remainder of this paper is organized as follows. Section 2 gives an overview of the theoretical background, with particular reference on how social interaction is addressed and measured. Section 3 describes the technical details of Sync'n'Move. Finally, evaluation results are discussed in Section 4.

2 Background

Our approach to the complex phenomenon of synchronization in social interaction is based on analysis of Phase Synchronization (PS) [8] In our model each user is represented with an n-dimensional state vector in which the n dimensions are n features like, for example, trajectories of body segments, amount of motion, audio descriptors and so on. On this assumption, interaction is addressed considering how the state vectors evolve together in time and extracting, from this joint dynamics, indices of a *global* system's behavior.

Phase synchronization can be measured using techniques based on the recurrence property of dynamical systems [9]. Recurrence Plots (RPs) [6] and Recurrence Quantitative Analysis (RQA) [7], [11] are techniques providing qualitative and quantitative information on systems' dynamics and their interrelations in terms of trajectories in a chosen features space. More specifically, a RP is a time-time binary colorimetric plot displaying all time instants in which recurrences in the state of a system are observed, whereas RQA quantifies the graphical patterns occuring in RP.

Let us consider two systems, identified by their state vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. RP for the first system (the same formula is also valid for the second system by replacing $\mathbf{x}$ with $\mathbf{y}$) is defined through the recurrence matrix:

$$R_{i,j}(\varepsilon) = \Theta\Big(\varepsilon - \parallel \mathbf{x}_i - \mathbf{x}_j \parallel\Big) \qquad \text{i}, j = 1...N \tag{1}$$

where, $\mathbf{x}_{i,j} \in \mathbb{R}^n$ are the system states at times i and j, N is the number of the states, ε is a closeness threshold, $\|\|$ and Θ are a norm (e.g., euclidean norm, minimum or maximum norms can be adopted) and the Heaviside function, respectively. The elements $R_{i,j}$ are binary values (0 or 1) according to whether $\mathbf{x}_i$ is *close* or not to $\mathbf{x}_j$.

By applying RQA to a RP we compute the probability $\hat{p}(\varepsilon, \tau)$ that the system recurs at a certain state after some time τ [10]. The estimate of this probability can be written as:

$$\hat{p}(\varepsilon,\tau) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} R_{i,i+\tau}(\varepsilon) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} \Theta(\varepsilon - \parallel \mathbf{x_i} - \mathbf{x}_{i+\tau} \parallel) \quad (2)$$

Again the same formula is valid for the second system by replacing $\mathbf{x}$ with $\mathbf{y}$. Finally we compute the PS between the two systems with the Correlation Probability of Recurrence (CPR), defined as:

$$PS = CPR = \langle \bar{p}_{\mathbf{x}}(\varepsilon,\tau)\bar{p}_{\mathbf{y}}(\varepsilon,\tau)\rangle \quad (3)$$

where $\bar{p}_{\mathbf{xy}}(\varepsilon,\tau)$ are the functions $\hat{p}(\varepsilon,\tau)$ normalized to zero mean and unitary standard deviation.

In the Sync'n'Move application we present in this paper, we aim to study the realtime synchronization between two systems: two users that move two mobile phones in space. In this case $\mathbf{x}$ and $\mathbf{y}$ are 1-dimensional state vectors representing the acceleration measured by on-board accelerometers the users mobile phones are endowed with. High phase synchronization between such accelerations corresponds to enhanced audio rendering as output, otherwise the audio output is minimal.

3 Application description

Our application is developed in the SAME networked platform for mobile, experience-centric, and context-aware active music listening. The SAME platform is an end-to-end framework (i.e., between clients of a mobile service, producers and consumers of content) for context-aware, experience-centric mobile music applications, enabling embodiment and control of music content by user behaviour. It facilitates rapid prototyping of context sensitive experience-centric embodied music applications and it enables intelligent, real-time, distributed processing of integrated music, video, and multimodal signals through a generalized plugin mechanism, aimed at open connectivity of mobile music systems and emerging media centers and industrial standards. The platform includes one or more servers running software environments, such as for example EyesWeb XMI [4] and MAX MSP [2]. Sync'n'Move runs on the SAME platform by allowing mobile phones to be connected and communicate to EyesWeb XMI.

Figure 1 sketches out how our application works. Two users freely move their mobile phones and their hand/body acceleration is detected and measured by the tri-axial accelerometer embedded in the mobile. An index of phase synchronization is extracted from their gesture: every time the users succeed in synchronizing (index is high) the music orchestration and rendering is enhanced; if instead users cannot synchronize (i.e., the phase synchronization index is low), the music gradually looses sections and rendering features.

Figure 1 shows the main modules composing our application: *the data acquisition module* and *the feature extraction & audio processing module.*

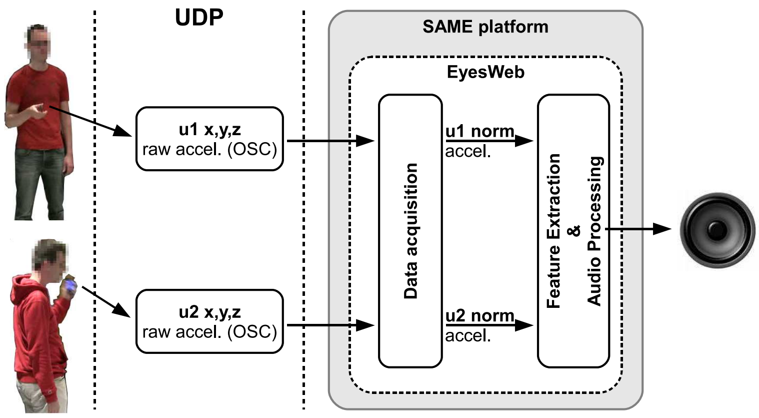

Fig. 1. Architecture of the Sync'n'Move application

3.1 The data acquisition module

This first module reported in Figure 2 acquires, calibrates, and computes the normalized acceleration captured by the mobile phones the users are moving. Each mobile runs a Python script that collects data from the accelerometer and creates an OSC packet in the form:

```
/synchronizer ax ay az
```

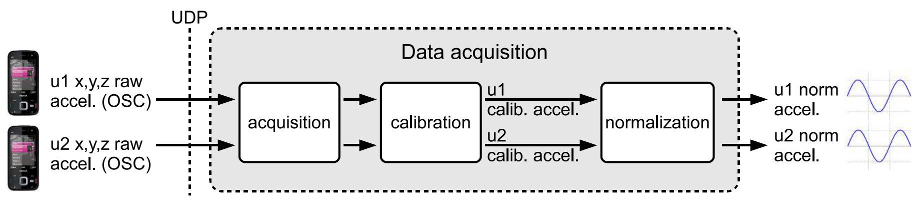

Fig. 2. Data acquisition module

The above packet is sent via UDP to the SAME platform running EyesWeb where the *raw accelerations* on the 3 axis are extracted from the OSC packet (*acquisition* box in Figure 2). The *calibration* and *normalization* blocks are necessary since every accelerometer has a different ground reference, that is, the *max* and *min* values of **g** change on every axis. Here we report the operations performed by the calibration block on the x axis, the same computation is necessary also on the other two axis:

$$A_{C_x} = A_{raw_x} - IRD_x; \quad IRD_x = \frac{gx^+ + gx^-}{2}; \tag{4}$$

in which:

– IRD_x is the *Inter Range Difference*, that is, the half of the difference between the maximum $\mathbf{g}$ measured on x^+ and x^-;
– A_{C_x} is the calibrated acceleration on x axis, that is, the acceleration obtained by subtracting the IRD_x from the raw acceleration on x, that is, A_{raw_x};

Finally the normalization block computes the absolute value of acceleration and nomalizes it in the range $[0, 1]$ subtracting $\mathbf{g}$ in order to ignore it:

$$A_{normalized} = \frac{\sqrt{A_{C_x}^2 + A_{C_y}^2 + A_{C_z}^2} - \mathbf{g}}{A_{MAX}}; \quad (5)$$

where A_{MAX} is the maximal absolute value of acceleration detected by the phone.

3.2 The feature extraction & audio processing module

Figure 3 shows the architecture of the feature extraction & audio processing module. This module is responsible for computing the phase synchronization index, which is used for controlling audio processing. From the normalized ac-

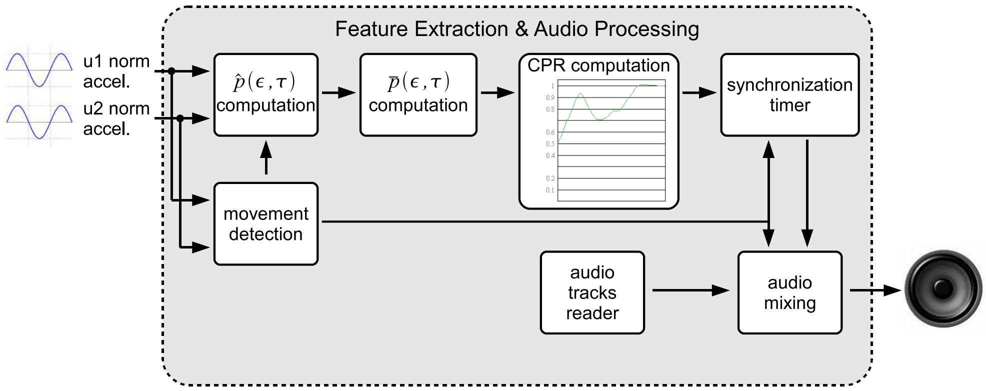

Fig. 3. The feature extraction & audio processing module

celerations, we compute the probabilities of recurrence $\hat{p_x}(\varepsilon, \tau)$ and $\hat{p_y}(\varepsilon, \tau)$ and we normalize them to obtain the probabilities $\bar{p_x}(\varepsilon, \tau)$ and $\bar{p_y}(\varepsilon, \tau)$ having zero mean and unitary standard deviation. The next step is the computation of CPR. The final output of the application is an audio content produced by the *audio processing* block. This content changes according to the synchronization degree between the users. The following three cases can occur:

– no audio: the users are not interacting at all, that is they are not moving their mobile phones. In this case the *movement detection* block detects that the two accelerations are equal to zero and inhibit audio generation;

– metronomic audio: (i) only one user is moving or (ii) both are moving but they are not synchronized. In the first condition, the *movement detection* block detects that just one of the accelerations is different from zero and enables the generation of a metronomic section in the audio output, e.g., the charleston instrument. In the second condition, the CPR is computed but it is too low to allow the generation of the full audio output.
– full audio: the two users are moving in a synchronized way. The CPR assumes a high value (almost one) and the *synchronization timer* measures the time along which the two users keep synchronized. According to the duration of this time, new sections are added to the audio content: the longer is the synchronization time the larger is the number of the enabled instruments, e.g., drums, bass and guitar, voice.

An example of how Sync'n'Move works can be found at:

`www.sameproject.eu/Demos`

4 Evaluation

Sync'n'Move was tested during the multidisciplinary encounter Agora Festival 2009 in Paris [3]. We gathered qualitative information on the application using an anonymous assessment questionnaire[1] filled up by 22 participants (19 male and 3 female from different european and extra-european countries) attending the Festival. Mean age of the participants was 33.7 years (range 18y-64y). All the participants were volunteered for this study and they were only asked to have a spontaneous behavior as much as possible. Before their performance, they are provided with a short demonstration of how the system works.

The participants were classified in two groups of users: expert and non-expert users. The experts'group was composed by 12 subjects having attested expertise in the musical field such as composers and professional musicians.

The questionnaire was composed by three parts: the first one, collecting general information about the participants such as gender, age, nationality and work; the second one, including general questions on the SAME applications presented at Agora Festival; the third one concerning the evaluation of each single application. This last part was different for each application. The analysis in this paper took into account data only from the first and the third part of the questionnaire. The third part of the questionnaire conceived for Sync'n'Move was composed by four questions about understanding, control, interaction level, fun, interest, future exploitation, engagement, and pleasure. Participants were asked to express their ratings on scales divided into eleven steps ranging from *not at all* to *very*. In the final part of the questionnaire, participants could write their comments and suggestions. We took into account partially filled up questionnaires also.

[1] The questionnaire was designed by the SAME consortium `http://www.sameproject.eu`.

Figure 4 summarizes the results for both groups simultaneously, whereas Figure 5 shows the results obtained for the expert users (Left Panel) and for the non-expert users (Right Panel) respectively. The y-axis shows the rates range expressed as numerical values (0 for *not at all*, 10 for *very*), the x-axis shows all the items from the questionnaire: *understanding*, *control*, *interaction*, *fun*, *interest*, *future*, *engagement*, *pleasure*.

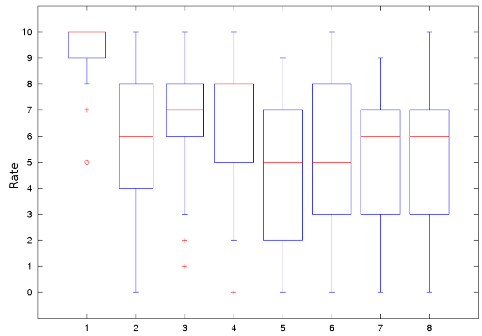

Fig. 4. The global box plot. The numbers on the x-axis stand for the items: 1-understanding, 2-control, 3-interaction, 4-fun, 5-interest, 6-future, 7-engagement, 8-pleasure.

From the inspection of the global box plots (Figure 4), we can infer that Sync'n'Move was very easy to understand (median=10). More specifically, non-expert users answered to the question about how the application works giving rates higher (full 10) than those given by expert users.

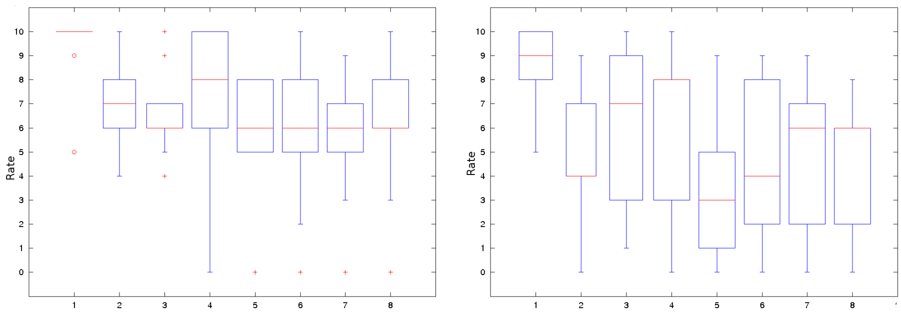

Fig. 5. Box plots of the rates of the non-expert users (Left Panel) and of the expert users (Right Panel).

Globally the rates given by non-expert users are higher than those reported by the experts. Non-expert users evaluated more than satisfactory the level of *control* and the *fun* during their experience. They gave neutral rates for *interaction*, *interest*, *future*, *engagement*, and *pleasure*.

Differently, the rates provided by expert users were more spread, for example the 50% of ranked rates for *interaction* ranges from 3 to 9, whereas the 50% of ranked rates for *engagement* ranges from 2 to 7. This may be due to the fact that expert users could perceive better than non-expert users that the audio tempo generated by Sync'n'Move was not matching in time with the tempo chosen by the users. This could also explain the low rates obtained for *control*, *interest*, and *future*. Moreover, Sync'n'move exploits very simple music content and interaction mechanisms, whereas expert users probably expect more complex paradigms and content.

5 Conclusion

This paper presented an application on active music listening. In Sync'n'Move two users generate an audio content by syncronizing their movements using mobiles phones as a collaborative interface. Evaluation shows that users are engaged in the application and that they well-understood how it works and the interaction paradigm we proposed.

We would like to improve the control that users have on generation of the music sections, e.g., enabling audio tempo to vary depending on the users' movement frequency, and allowing users to play their preferred music in Sync'n'Move. In this last case, we plan to use an audio sections separation algorithm developed within the SAME project.

Acknowledgments We thank Paolo Coletta, Alberto Massari, and Corrado Canepa. We also thank Carlo Chiorri for his precious help in the set up of the questionnaire.
This work is partially supported by the FP7 EU ICT SAME Project n. 215749 on active music listening (www.sameproject.eu).

References

1. I. Laso-Ballesteros and P. Daras (Eds.), User Centric Future Media Internet, EU Commission, September 2008.
2. `http://www.cycling74.com`.
3. `http://agora2009.ircam.fr/843.html`.
4. A. Camurri, P. Coletta, G. Varni, and S. Ghisio. Developing multimodal interactive systems with EyesWeb XMI. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 305–308. ACM New York, NY, USA, 2007.
5. A. Camurri and G. Volpe. Active and personalized experience of sound and music content. In *Proc. 12th IEEE International Symposium on Consumer Electronics*. IEEE Press, April 2008.

6. J. P. Eckmann, S. O. Kamphorst, and D. Ruelle. Recurrence plots of dynamical system. *Europhysics Letters*, 5:973–977, 1987.
7. N.Marwan, M. C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438:237–329, 2007.
8. A. Pikovsky, M.G. Rosenblum, and J. Kurths. *Synchronisation: a Universal Concept in Nonlinear Sciences*, volume 12 of *Cambridge Nonlinear Science Series*. Cambridge University Press, Cambridge, 2001.
9. H. Poincaré. Sur la probleme des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–271, 1890.
10. M.C. Romano, M. Thiel, J. Kurths, I.Z. Kiss, and J.L. Hudson. Detection of synchronisation for non-phase coherent and non-stationarity data. *Europhysics Letters*, 71(3):466–472, 2005.
11. J. Zbilut and C. L. Webber Jr. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 5:199–203, 1992.