Embodied cooperation using mobile devices: presenting and evaluating the Sync4All application

Giovanna Varni
InfoMus-DIST-University of Genoa
Viale Causa 13
16145 Genoa, Italy
giovanna@infomus.org

Maurizio Mancini
InfoMus-DIST-University of Genoa
Viale Causa 13
16145 Genoa, Italy
maurizio.mancini@unige.it

Gualtiero Volpe
InfoMus-DIST-University of Genoa
Viale Causa 13
16145 Genoa, Italy
gualtiero.volpe@unige.it

ABSTRACT

Embodied cooperation "arises when two co-present, individuals in motion coordinate their goal-directed actions". The adoption of the embodied cooperation paradigm for the development of embodied and social multimedia systems opens new perspectives for future *User Centric Media*. Systems for embodied music listening, which enable users to influence music in real-time by movement and gesture, can greatly benefit from the embodied cooperation paradigm. This paper presents the design and the evaluation of an application, *Sync4All*, based on such a paradigm, allowing users to experience social embodied music listening. Each user rhythmically and freely moves a mobile phone trying to synchronise her movements with those of the other ones. The level of such a synchronisation influences the music experience. The evaluation of Sync4All was aimed at finding out which is the overall attitude of the users towards the application, and how the participants perceived embodied cooperation and music embodiment.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Representation (HCI)**]: User Interfaces—*User-centered design*

General Terms

Human Factors, Design

Keywords

Embodied cooperation, nonverbal social behaviour, mobile music applications

1. INTRODUCTION

Embodied cooperation "arises when two co-present, individuals in motion coordinate their goal-directed actions". [20]. In contrast with traditional game-theory approaches that do not require individuals to have bodies or to be able to move and detect information about another's movement, but rather focus on top-down cognitive judgements, the embodied cooperation paradigm requires that users act and interact in a real environment by means of features that are detectable by the perceptual system. As Marsh et al. argue: "embodied cooperation was motivated by the intuition that the presence of another person extends the action affordances that are possible for the individual" [20]. This and related topics were addressed in several research fields such as psychology and cognitive science (e.g., [28][31]). At the present, the trend of the game industry (e.g., Nintendo and Microsoft) shows that the embodied cooperation paradigm is successful and more and more popular to engage users in games (e.g., [1][19]). Embodied companions offer opportunities for increasing engagement, coordination, and interaction, and for studying how basic abilities of cooperation can be acquired via social learning. However, the exploitation of embodied cooperation for the development of social machines is still under-investigated. Only in the past few years the HCI research community manifested an increasing interest on this topic and most of the existing multimedia interactive systems and Internet applications, such as social networks and search engines, are still intended for a single user.

The embodied cooperation paradigm opens new perspectives for future *User Centric Media*, in which users' quality of experience is enhanced because they become active members of the overall media chain by generating, distributing, and experiencing high-quality media content [17]. In such a framework embodiment is a key factor. For example, systems for embodied music listening [35] create a new kind of technology-mediated experience of sound and music, where users can influence in real-time the music they are listening to by movement and gesture.

This paper presents the design and the evaluation of an application, *Sync4All*. The goal of the application is to listen to music in a different way, with respect to the current "passive" paradigm: the listening experience becomes an active exploration of the music content carried out collaboratively by body movement. In Sync4All each user rhythmically and freely moves her mobile phone trying to synchronise with the other users. Synchronisation is measured as *Phase Synchronisation* of gestures, a consolidated coordination metric [29][2]. The possibility of establishing coordination is deemed as a cue for eliciting cooperation [16].

Sync4All was developed in the framework of the EU-ICT

SAME Project (www.sameproject.eu) and is a first example of social embodied music listening as targeted in future User Centric Media.

The evaluation of Sync4All aimed at finding out which is the overall attitude of the users to the application, and how the participants perceived embodied cooperation and music embodiment.

2. RELATED WORK

In the specific field of User Centric Media, and in particular of embodied music systems, research on collaborative systems mainly concerned music making. Blaine and Fels [3] claim that musical collaboration systems commonly restrict musical control, facilitating novices' participation in the musical experience. According to Jordà [13], multi-user instruments facilitate responsiveness and interaction between each performer and the instrument, and also between performers. Several systems were developed for collaborative music making. Just to mention some of them, the ReacTable [14] allows a group of people to share control of a modular synthesiser by manipulating physical objects on a round table; Audiopad is a composition and performance instrument for collaborative electronic music which tracks the positions of objects on a tabletop surface and converts their motion into music [26]; TinyTune is a collaborative musical instrument using sensor motes [24]; JamMo is a mobile technological tool for music making for young children [22].

Whereas collaborative music making addresses systems enabling a group of users to create and play a (new) music piece, collaborative music listening concerns systems where a group of users can cooperate in listening to and possibly modifying an already existing and pre-recorded music piece. Camurri [7] proposed an early pioneering system where the user full-body rhythmic movements were analysed in real-time and compared with the beat of a song (extracted from the MIDI music signal). Leman et al. [18] reworked the concept of social music game: the movement beat of multiple users was extracted and compared with the beat of the music the users were listening to. Users could compete among them or collaborate to win the game. Stokholm and Pasquier [32] implemented a system mixing audio representations of the mood of several users to increase collaboration and empathy among users. Vinyes and colleagues developed the Audio Explorer system, enabling users to concurrently modify the audio mixing of a piece of music downloaded from the Web and to share the resulting content [34]. Mappe per Affetti Erranti [35] allowed the exploration, by means of the movement of multiple users, of multiple layered physical and affective maps enabling the users to influence both the polyphonic structure of a music piece and its expressive interpretation. Whereas, on the one hand, these systems witness the relevance of embodiment and social interaction for novel music listening applications, on the other hand most of them do not address embodied cooperation explicitly. That is, these systems do not directly use embodied cooperation metrics to influence music content listening.

Systems that explicitly measure social signals and perform analysis of social interaction are emerging in the HCI community (e.g., see [33] for a survey on social signal processing), but are still missing in the music scenario. For example, Pentland and colleagues investigated the emergence of formal and informal roles in the framework of talk-shows, movies, and meetings [27]; Hung and colleagues studied dominance and group cohesion in small group meetings [12][11]; the EU-ICT Network of Excellence SSPNet (http://sspnet.eu) is centred on the analysis of social signals with a special focus on nonverbal behavioural cues. Research focusing on the embodiment of social interaction includes studies on social robotics (e.g., [9][6]) and the prototyping of robot companions. For instance, since long time Breazeal and colleagues investigated several approaches to build socially intelligent robots (e.g.,[5]); among the many examples of social robots, the Walk-Mate robot [21] is a virtual locomotion collaborative walking system, able to support the walking of Parkison's disease and hemiplegia patients, that explicitly exploits analysis of interpersonal synchronisation. Further studies addressed automatic analysis of face-to-face conversational interaction for application to embodied conversational agents (e.g., see [15]).

Grounding on current research on social signal processing, Sync4All is an example of application for social embodied music listening that adopts the embodied cooperation paradigm and that explicitly relies on the extraction and processing of social signals, namely synchronisation.

3. DETECTING SYNCHRONISATION IN A GROUP

3.1 Interaction Scenario

The interaction scenario of Sync4All consists of four (or more) users holding a mobile phone in their hands. The users may meet either at home or in a public space (e.g., a disco, a pub) and may choose a pre-recorded multi-track music piece they want to actively explore. In order to explore the piece, they are simply asked to freely move the mobile, for example by dancing, by shaking it, or by using it like a baton to conduct an orchestra. As soon as one of the users starts moving her mobile a rhythmic hi-hat sound (part of the drums section) is played. The user can choose to follow this rhythm and synchronise her gesture with it or to perform other rhythmic patterns. Then, if other users want to join to the collaborative listening, they have to move their mobile phones trying to reach synchronisation at the rhythmic level: that is, they can perform any kind of gesture provided that it is in time with the rhythm of the first one. As far as a pair of users gets synchronised over a certain threshold, another section of the piece emerges and, according to the increase in the synchronisation, progressively adds to the rhythmic pattern. Each pair of mobile phones (i.e., users) is associated to a specific music section. By synchronising in different pairs, the users may explore and listen to several different sections of the piece. While larger subgroups synchronise (e.g., a subgroup of 3 users), the music sections associated with all the pairs in the subgroup are reproduced together, making the music output richer. Thus, the more the users are able to synchronise as a group, the more they can listen to sections of the music. Only when all the users synchronise as a whole single group, the music piece is reproduced in all its sections. So the users need to cooperate as a group to fully reconstruct the piece.

3.2 Application design

Each user connects her mobile via wireless network to a server running Sync4All: the set of the connected users forms a dynamically changing graph whose vertices are the users themselves. Each mobile is endowed with on-board accelerometers. Once the connection is successfully established, each mobile sends the detected accelerations to the server. It evaluates which pairs of users are synchronised and the strength of such a synchronisation. The number of synchronised pairs and the extent to which they are synchronised contributes to define the graph topology: an edge between two nodes appears every time a pair reaches a remarkable synchronisation. The weight of the edge depends on the strength of the synchronisation. Each edge is coupled with a music track, and the corresponding weight is mapped onto the volume of the track. For example, in the graph in Figure 2, three pairs out of six are synchronised and the corresponding music tracks are reproduced. Only when the graph is topologically connected (that is all the pairs are synchronised), the audio output includes all the tracks. Once a user disconnects her mobile from the server, the corresponding vertex is pruned from the graph.

In this study Sync4All was implemented and evaluated on four users. Figure 1 depicts the architecture of the system:

- *calibration*: acceleration data from the users' mobile phones is processed to filter out gravity acceleration;
- *probability computation*: for each user the probability of recurrence of the acceleration time series is computed [25];
- *PARS computation*: synchronisation between each pair of users is estimated by using techniques from [25], [30];
- *audio mixing*: the audio output is reproduced.

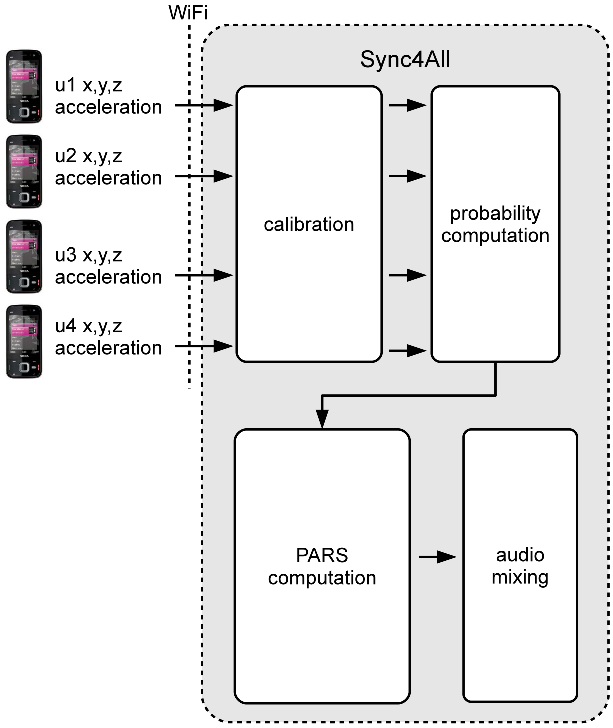

Figure 1: The architecture of Sync4All. Acceleration data from mobile phones is processed to compute synchronisation between each pair of users. Synchronisation controls audio mixing.

3.2.1 Calibration

3D (x, y, z) acceleration data captured by the users' mobiles accelerometers is received via UDP. On every axis, accelerometers detect accelerations in a range varying in $[0, A_{MAX}]$. Moreover, gravity acceleration is also included in the measurement. To obtain absolute calibrated acceleration, that is, a single value varying in $(0, 1)$ the following computation is performed:

$$A = \frac{\sqrt{A_x^2 + A_y^2 + A_z^2} - g}{A_{MAX}}; \qquad (1)$$

where: A_x, A_y, A_z are the components of the detected acceleration along the three axes, and g is the magnitude of gravity acceleration.

3.2.2 Probability computation

This module computes the normalised probability of recurrence $(\hat{p}(\epsilon, \tau))$ of the acceleration for each user. That is, the probability that each user's acceleration will recur for a given value after a time lag τ. Given the time series A_i of the acceleration of one user, the probability of recurrence is computed by using:

$$\hat{p}(\varepsilon, \tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} R_{i,i+\tau}(\varepsilon) =$$

$$= \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \Theta(\varepsilon - \parallel A_i - A_{i+\tau} \parallel) \qquad (2)$$

where ϵ is a fixed threshold, defining the neighbourhood in which two samples are considered as coincident, and Θ is the Heaviside function. Normalisation is such that the mean probability of recurrence is equal to 0 and its standard deviation is equal to 1.

3.2.3 PARS computation

Dealing with oscillatory signals, the $\hat{p}$-s for each signal show maxima for some τ. When the signals are synchronised these maxima coincide. In Sync4All the signals are the users' calibrated accelerations (A). A measure of the coincidence for each pair of accelerations is given by:

$$CPR_{a,b} = \langle \hat{p}_a(\epsilon, \tau)\ \hat{p}_b(\epsilon, \tau) \rangle \quad CPR_{a,b} \in [0, 1] \qquad (3)$$

where a and b refer to the two accelerations in the pair, $\hat{p}_i$ is the normalised probability of recurrence as described above, and $\langle , \rangle$ indicates the correlation operation.

CPR provides information about synchronisation in a pair of users, but is not able to distinguish direct from indirect interactions in a network of users. To address this issue *partial recurrence based synchronisation (PARS)* [23] is conceived. For each pair (a, b) the $CPR_{a,b}$ index is computed and all the computed indices are arranged in a matrix. The resulting matrix is a symmetric matrix with the elements on the main diagonal equal to 1. For instance, with $K = 4$ users (a, b, c, d) the resulting matrix is:

$$\hat{\mathbf{P}} = \begin{pmatrix} 1 & \hat{p}_{a,b} & \hat{p}_{a,c} & \hat{p}_{a,d} \\ \hat{p}_{b,a} & 1 & \hat{p}_{b,c} & \hat{p}_{b,d} \\ \hat{p}_{c,a} & \hat{p}_{c,b} & 1 & \hat{p}_{c,d} \\ \hat{p}_{d,a} & \hat{p}_{d,b} & \hat{p}_{d,c} & 1 \end{pmatrix}$$

Then, this matrix is inverted and the PARS indices are defined as:

$$\hat{p}_{i,j|h,k} = \frac{|\ \hat{p}^{-1}_{i,j}\ |}{\sqrt{\hat{p}^{-1}_{i,i}\hat{p}^{-1}_{j,j}}} \quad (4)$$

Each of these elements quantifies the synchronisation between the i^{th} and j^{th} users filtering out the h^{th} and k^{th} users.

3.2.4 *Audio mixing*

The output of Sync4All is an audio content that varies depending on the synchronisation of the users' movement. Audio is controlled by dynamically computing the mixing matrix using the elements $\hat{p}_{i,j|h,k}$ to determine which audio tracks are activated and their volume. Several situations may occur:

- no audio: the users are not interacting at all, that is they are not moving their mobile phones. In this case the audio reproduction is inhibited;
- metronomic audio: (i) only one user is moving or (ii) more users are moving but they are not synchronised. In such a situation the application output is a percussive audio only (for example, drum hi-hat).
- partial audio: users are moving and some of them are synchronised in pairs. According to which pairs are synchronised the application enables the reproduction of one or more audio tracks: e.g., the guitar, the bass, the keyboard, and so on.
- full audio: as the number of synchronised user pairs is over a given threshold the application enables all the audio tracks, including a further track, for example a singer's voice.

For example, in the graph of Figure 2 three pairs of users are synchronised, enabling the corresponding audio tracks.

Sync4All was implemented with the EyesWeb XMI framework (www.eyesweb.org). A video of the application is available at: `http://youtu.be/AGyunWTPTdg`.

4. EVALUATION

Sync4All was presented and evaluated during the public exhibition "Festival della Scienza" hosted in Genova (Italy) on November 2010. At this exhibition eight more sample mobile music applications, developed by the SAME Consortium, were demonstrated and evaluated (e.g., the Mobile Orchestra Explorer, see [10]). The evaluation of Sync4All was carried out both via an anonymous assessment questionnaire (partially used also for evaluating the other applications) and via time-series analysis of the acceleration data. The analysis of the data from the questionnaire, which is addressed in this paper, was aimed at finding out which is the overall attitude of the users to the application, and how the participants perceived embodied cooperation and music embodiment. The time-series analysis, which is not addressed here, will measure whether the application facilitates social interaction (i.e., the difference between what the users perceived and what they actually did). The questionnaire was composed by three parts:

- the first part was conceived to gain information about the habits of the participants in the daily use of the mobile phone. This information was used to create a profile of the users and to cluster them;

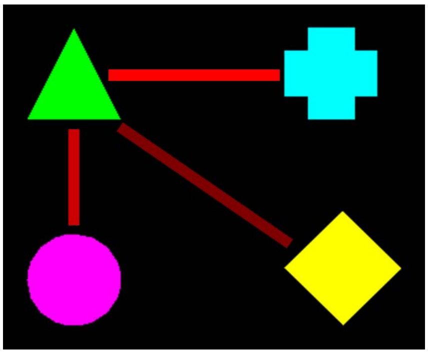

Figure 2: The graph showing the participants (geometric symbols) and their synchronisation (red edges). The transparency of each edge is in direct ratio to the strength of the synchronisation.

- the second part was conceived to measure the attitude of the participants to the application;
- the third part addressed how the participants perceived embodied cooperation and music embodiment.

A blank section was left at the end of the questionnaire to gather the comments and the suggestions from the participants. The questionnaire is reported in Appendix A.

4.1 Method

4.1.1 *Participants*

Seventy-two individuals tested the application as volunteers. The questionnaire was filled up by 70 participants (38 male and 32 female) coming both from European (60 people from Italy, 4 from France, 1 from Serbia) and extra-European countries (1 from Algeria, 1 from Brazil, and 1 from Ecuador). Mean age was 22.9y (std=14.1y), the minimum age was 8y, whereas the maximum age was 56y. Most participants (67.1%) were students whose the main field of study was science. All but 4 participants (3 children and 1 young man) owned a mobile phone.

Authorisation was requested for children.

4.1.2 *Procedure*

The participants filled up the first part of the questionnaire and they had to carefully read a sheet on the purposes and the instructions explaining the experiment. They were asked: (i) to synchronise, by moving their mobile and (ii) to "listen" to the music tracks all together as long as possible. Before the beginning of the recordings, the participants were allowed a short training (about 1 minute) over which they could try the application to reach synchronisation and try different gestures. Along this time, a visual rendering of the graph was displayed in real-time on a screen to help the participants to understand how the application works. Figure 2 shows a screenshot of the graph: each participant is portrayed as a geometric symbol, the transparency of the edges is in direct ratio to the strength of synchronisation.

After this training, the participants tested the application over 4 minutes without any visual feedback, and then they filled up the second and the third part of the assessment questionnaire. The overall duration of the experiment was 5 minutes.

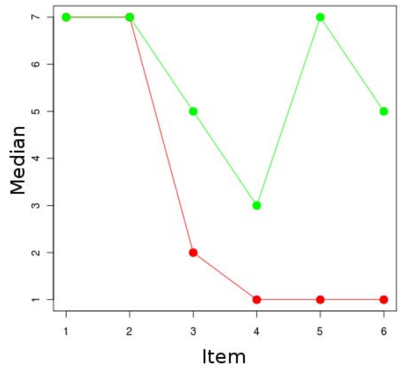

Figure 3: Item-by-item medians of the scores rated by advanced (green line) and basic (red line) users. The labels on the x-axis stand for: 1: calling, 2: sending SMSs, 3: taking pictures, 4: recording videos, 5: listening to music, 6: playing games and applications. The y-axis shows the seven points of the Likert scale: 1: never, 2: less than once a month, 3: once a month, 4: once a week, 5: sometimes a week, 6: once a day, 7: sometimes a day.

4.1.3 Participants' characterisation

Participants were clustered depending on their skill in using mobile phones daily. The data set for this analysis came from the first part of the questionnaire collecting information on the habits of the participants about mobile phones. A 7-points Likert scale was set up to profile users. The Likert items addressed how long participants spend in calling, sending SMSs, taking pictures, recording videos, listening to music, and playing games or other applications. The answers ranged from *never* to *several times a day*. Four participants chose to not fill up this part of the questionnaire. A BIC value-based Xmeans algorithm [8] was applied to this data set, and two clusters were identified and labelled as advanced users' cluster and basic users' cluster. The former cluster was composed by 30 (45%) participants, the latter by 36 (55%) participants. The basic users' cluster included participants with mean age equal to 17.8y (std=7.9y), whereas the advanced users' cluster mean age was 18.2y (std=7.9y). The 65% of the participants having some expertise in music (playing music or using computer music tools) belongs to the advanced users' cluster. The same cluster included the 51.6% of participants practising a regular physical activity. Figure 3 depicts the item-by-item medians of the scores rated by advanced and basic users.

A Wilcoxon-Mann-Whitney test was run to compare the scores from these two clusters. A Bonferroni correction of the statistical significance level was needed to take into account the multiple comparisons. The corrected level of statistical significance was $p=0.01$. The null-hypothesis was rejected for the items related to the use of the more advanced functionalities of the mobile phones, that is taking pictures ($p<0.01$, $r=0.71$), recording videos ($p<0.01$, $r=0.71$), listening to music, playing games ($p<0.01$, $r=0.91$).

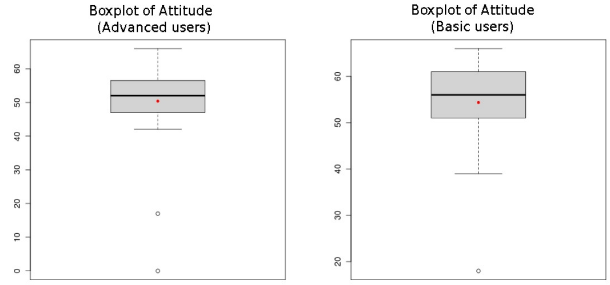

Figure 4: The boxplots of the participants attitude to Sync4All (left panel: advanced users, right panel: basic users). The y-axis shows the summed up scores (ranging from 0 to 66), mean is depicted by a red dot. Possible outliers are marked with empty dots.

4.2 Results

4.2.1 Users' attitude to Sync4All

A 11-points Likert scale including 6 items was administered to the participants in order to measure their attitude to the application. These items composed the second part of the questionnaire. The scores ranged from 1 to 11 for each item. Questions concerned the following: understanding, playability, pleasure, interest, engagement, and forward-looking. The scores each participant rated on this scale were summed over all the items. Missing values were dealt with a pairwise deletion method. Figure 4 shows the boxplots of the attitude of the participants, both for the advanced and the basic users. Cronbach's alpha was computed to assess reliability; for both clusters alpha was greater than 0.7 (basic users'α=0.84, advanced users'α=0.93) assuring the reliability of the measure. The medians of the basic and the advanced users' clusters were 56 and 52, respectively. A Wilcoxon-Mann-Whitney test revealed that no significant difference can be claimed between these two clusters in terms of the attitude to the application ($p>0.05$). Although no significant difference emerged between the clusters, the high median values confirm that the participants' opinion on Sync4All was very satisfactory.

Overall, the users would advise their friends the application (only 2 users and 1 user of the basic and the advanced clusters expressed a negative opinion, respectively) and they would use it if the application were available on their mobile (basic users' cluster: 71.4% and advanced users' cluster: 91.6%, one user for each cluster did not answer).

4.2.2 Perception of embodied cooperation

Likert item 4.5 was related to perception of embodied cooperation; such an item was formulated according to previous studies for annotating social behaviour in a group [11], [4]. All participants but two (one for each cluster) answered this question: 93.1% of the basic users and 88.8% of the advanced users replied in the positive. Further, basic and advanced users show a median equal to 8 and 9, respectively. A Wilcoxon-Mann-Whitney test carried out on the scores did not reveal any difference in the scores ($p>0.05$). This is an interesting result because one would expect that advanced users are more acquainted with handling the mobile phone and then more inclined to embody it than basic users. However, the participants were asked to use the mobile in

Table 1: Comments and suggestions about Sync4All provided by the participants

# Participant	Comments and Suggestions
13, 133, 40, 65	Very interesting; beautiful.
105, 73	It was very difficult.
107	I prefer the true musical instruments, but it is very interesting as group performance.
184, 183,132	I suggest to match every person with one instrument and to handle synchronisation in a different way, so that people can create music.
173	Good for professional market, not for home use.
131	Good application, I suggest a larger playlist.
132	Funny and agreeable application.
130	I suggest to modify the application by allowing the participants to perform more free gestures. This will make the application funnier and less mechanical.
32	Good for working as a group.
20	Very cool! I would like to install this application when I will have my own mobile phone.
67	This application can be used by young people only.
164	I suggest to synchronise the tempo of the music with the tempo of the gestures.

an unusual way; this could explain the obtained result.

4.2.3 Embodied cooperation and music listening

The final Likert item of the questionnaire addressed the music embodiment issue. The medians of the scores of the two clusters were 11 (basic users) and 9.5 (advanced users). A Wilcoxon-Mann-Whitney test was run to evaluate the difference in the answers the participants provided: basic and advanced users were aware at a similar level that music changed according to the level of group cooperation ($p>0.05$).

4.2.4 Comments and suggestions

The questionnaire included a blank section in which the participants could freely quote their comments and suggestions. Comments are reported in Table 1. Most of them were positive and the suggestions highlighted very important issues to be taken into account in the future. However, maybe due to the long time the users spent for visiting the whole exhibition and to perform the evaluation of other applications, only 27.14% of the participants filled up this section.

4.3 Discussion

The above results suggest that the design of future interactive embodied music listening applications could be positively welcomed by both advanced and basic mobile devices users. By collaborating together, users will be able to explore, shape and perceive music content, an experience that nowadays is reserved only to people playing music.

5. CONCLUSION

This paper presented the design and the evaluation of Sync4All, an application based on the embodied cooperation paradigm and aimed at promoting the emergence of social embodied music listening. Evaluation showed that both advanced and basic users (i) have a very positive attitude to Sync4All, (ii) perceived the embodied cooperation (iii) were aware that music changed according to the level of cooperation in the group. Free comments and suggestions further confirmed that the participants' opinion on Sync4All was very satisfactory.

In the near future, users will be allowed to modify, by their gesture, the tempo of the music piece resulting from their collaborative interaction. Also, the application's response time, that is, the delay between the moment in which the users starts to move and the one in which the music piece is modified/rendered, will be improved.

6. ACKNOWLEDGMENTS

This work was partially supported by the EU-ICT SAME Project (2008-2010). The authors thank Antonio Camurri, Donald Glowinski, Paolo Coletta, and Corrado Canepa.

7. REFERENCES

[1] N. Bianchi-Berthouze, W. W. Kim, and D. Patel. Does body movement engage you more in digital game play? and why? In *ACII*, pages 102–113, 2007.

[2] G. Bingham, R. Schmidt, M. Turvey, and L. Rosenblum. Task dynamics and resource dynamics in the assembly of coordinated rhythmic activity. *Journal of Experimental Psychology: Human Perception and Performance*, 17:359–381, 1991.

[3] T. Blaine and S. Fels. Collaborative musical experiences for novices. *Journal of New Music Research*, 32:411–428, 2003.

[4] L. Braaten. Group cohesion: A new multidimensional model. *Group*, 15:39–55, 1991.

[5] C. Breazeal. *Designing sociable robots.* AAAI Press, 2002.

[6] C. Breazeal, A. Takanishi, and T. Kobayashi. Social robots that interact with people. In *Springer Handbook of Robotics*, pages 1349–1369. 2008.

[7] A. Camurri. Interactive dance/music systems. In *Proceedings Intl. Computer Music Conference (ICMC-95)*, pages 245–252, 1995.

[8] A. M. Dan Pelleg. X-means: Extending k-means with efficient estimation of the number of clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 727–734, San Francisco, 2000. Morgan Kaufmann.

[9] K. Dautenhahn. Socially intelligent robots: dimensions of humanŰrobot interaction. *Philosophical Transactions of the Royal Society B*, 362:679–704, 2007.

[10] D. Glowinski, M. Mancini, and A. Massari. Evaluation of the mobile orchestra explorer paradigm. In *Proceedings of the 4th International ICST Conference on Interactive Entertainment.* LNICST, Springer, 2012.

[11] H. Hung and D. Gatica-Perez. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6):563–575, 2010.

[12] H. Hung, Y. Huang, G. Friedland, and D. Gatica-Perez. Estimating dominance in multi-party

meetings using speaker diarization. *IEEE Transactions on Audio, Speech & Language Processing*, 19(4):847–860, 2011.
[13] S. Jordà. Multi-user instruments: models, examples and promises. In *Proceedings of the 2005 conference on New interfaces for musical expression*, NIME '05, pages 23–26, Singapore, Singapore, 2005. National University of Singapore.
[14] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 139–146, New York, NY, USA, 2007. ACM.
[15] S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, 52(6):587–597, 2010.
[16] R. Kurzban. The social psychophysics of cooperation: nonverbal communication in a public good game. *Journal of Nonverbal Behavior*, 25:241–259, 2001.
[17] I. Laso-Ballesteros and P. D. (Eds.). User Centric Future Media Internet, EU Commission, September 2008.
[18] M. Leman, M. Demey, M. Lesaffre, L. van Noorden, and D. Moelants. Concepts, technology and assessment of the social music game 'sync-in team'. In *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering*, 2009.
[19] S. E. Lindley, J. L. Couteur, and N. Berthouze. Stirring up experience through movement in game play: effects on engagement and social behaviour. In *CHI*, pages 511–514, 2008.
[20] K. Marsh, M. Richardson, and R. Schmidt. Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, 1:320–339, 2009.
[21] Y. Miyake. Interpersonal synchronization of body motion and the walk-mate walking support robot. *Robotics, IEEE Transactions on*, 25(3):638 –644, june 2009.
[22] M. Myllykoski and P. Paananen. Towards new social dimensions for children's music making - jammo as a collaborative and communal m-learning environment. In *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, ESCOM 2009, pages 366–371, 2009.
[23] J. Nawrath, M. C. Romano, M. Thiel, I. Z. Kiss, M. Wickramasinghe, J. Timmer, J. Kurths, and B. Schelter. Distinguishing Direct from Indirect Interactions in Oscillatory Networks with Multiple Time Scales. *Physical Review Letters*, 104(3):038701+, 2010.
[24] B. Newman, J. Sanders, R. Hughes, and R. Jurdak. Tinytune, a collaborative sensor network musical instrument. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 375–376, New York, NY, USA, 2010. ACM.
[25] N.Marwan, M. C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438:237–329, 2007.
[26] J. Patten, B. Recht, and H. Ishii. Audiopad: a tag-based interface for musical performance. In *Proceedings of the 2002 conference on New interfaces for musical expression*, NIME '02, pages 1–6, Singapore, Singapore, 2002. National University of Singapore.
[27] A. Pentland. Social signal processing. *IEEE Signal Proc. Mag.*, 2(4):108–111, 2007.
[28] M. Richardson, K. Marsh, R. Isenhower, J. Goodman, and R. Schmidt. Rocking together: Dynamics of unintentional and intentional interpersonal coordination. *Human Movement Science*, 26:867–891, 2007.
[29] M. Richardson, K. Marsh, and R. Schmidt. Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31:62–79, 2005.
[30] M. Romano, M. Thiel, J. Kurths, I. Kiss, and J. Hudson. Detection of synchronisation for non-phase coherent and non-stationarity data. *Europhysics Letters*, 71(3):466–472, 2005.
[31] R. Schmidt and B. O'Brien. Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9:189–206, 1997.
[32] J. Stockholm and P. Pasquier. Reinforcement learning of listener response for mood classification of audio. In *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering*, 2009.
[33] A. Vinciarelli, M. Pantic, H. Bourland, and A. Pentland. Social signals, their function, and automatic analysis: a survey. In *Proceedings of the 10th International Conference on Multimodal Interfaces, ICMI08*, pages 61–68, 2008.
[34] M. Vinyes, J. Bonada, and A. Loscos. Demixing commercial music productions via human-assisted time-frequency masking. In *Proceedings of the 120th AES Convention*, 2006.
[35] G. Volpe and A. Camurri. A system for embodied social active listening to sound and music content. *ACM Journal on Computing and Cultural Heritage*, 4:2:1–2:23, 2011.

APPENDIX

The questionnaire participants were asked to fill up is reported below. Only the parts of the questionnaire used for the analysis presented in this paper are included. The full questionnaire is available on the EU-ICT SAME project website.

PART I

1.1 Gender (F/M)
1.2 Age
1.3 Nationality
1.4 Occupation

2.1 Do you own a mobile phone? (No/Yes)
2.2 How often are you using the services of your mobile? (one answer among: never; less than once a month; once a week; several times a week; once a day; several times a day)

2.2.1 Making calls
2.2.2 Sending/receiving SMS messages
2.2.3 Taking pictures
2.2.4 Recording videos
2.2.5 Listening to music
2.2.6 Playing games
2.2.7 Other musical applications

3.1 Do you play a musical instrument? (No/Yes)
3.2 Do you make use of computer music technology tools? (No/Yes)
3.5 I listen to music (one answer among: never; less than once a month; once a month; once a week; several times a week; once a day; several times a day)
3.10 Do you regularly practice a physical activity? (No/Yes)

PART II

4.1 How would you assess the application? (11-step Likert items)

4.1.1 From *Very difficult to understand* to *Very easy to understand*.
4.1.2 From *Very difficult to play* to *Very easy to play*.
4.1.3 From *I did not enjoy it* to *I enjoyed it*.
4.1.4 From *Not interesting* to *Interesting*.
4.1.5 From *Not engaging* to *Engaging*.
4.1.6 From *Nothing for the future* to *Something for the future*.

4.2 Would you recommend this application to a friend? (No/Yes)
4.3 Would you use this application if it was available in your mobile phone? (No/Yes)

PART III

4.5 You were involved in the application with other people. Did you feel to interact with this? (No/Yes)

4.5.1 If Yes, how much? (11-step Likert item, from *Not very much* to *Very much*).

4.7 Were you aware that the music you heard was changing depending on your synchronisation with the other three people? (11-step Likert item, from *Not at all aware* to *Completely aware*).

This part finished with a blank space for collecting comments and suggestions for improvements.