A qualitative evaluation of augmented human-human interaction in mobile group improvisation

Roberto Pugliese*
Department of Media,
also Media Technology
Aalto University
School of Arts, Design and
Architecture also,
School of Science

Koray Tahiroğlu*
Callum Goddard*
James Nesfield*
Department of Media
Aalto University
School of Arts, Design and
Architecture

ABSTRACT

In this paper strategies for augmenting the social dimension of collaborative music making, in particular in the form of bodily and situated interaction are presented. Mobile instruments are extended by means of relational descriptors democratically controlled by the group and mapped to sound parameters. A qualitative evaluation approach is described and a user test with participants playing in groups of three conducted. The results of the analysis show core-categories such as ***familiarity with instrument and situation***, ***shift of focus in activity***, ***family of interactions*** and different ***categories of the experience*** emerging from the interviews. Our evaluation shows the suitability of our approach but also the need for iterating on our design on the basis of the perspectives brought forth by the users. This latter observation confirms the importance of conducting a thorough interview session followed by data analysis on the line of grounded theory.

Keywords

Collaborative music making, evaluation methods, mobile music, human-human interaction.

1. INTRODUCTION

Activities associated with music making, such as non-verbal, gesture driven, extra-musical communication are natural ingredients of group improvisation. These activities have been part of the musical interaction along with the exchanged musical events in the moment of playing. Mobile phone technology can support collective, mobile and casual music-playing by providing a multitude of sensors that can facilitate interactive performance [4]. In this paper we intend to show that novel strategies in the development of new interfaces with extended systems enable musicians to become more aware of their social interaction. The main emphasis of our current research is to understand more about bodily interaction among participants, specially in relation to their location, distance and coordination as a group, and using these parameters as sound producing events. The research, *The Notion of Participatory and Enacting Sonic Interaction* (PESI), deals with the social aspect of participation and musicianship dynamics in collective music performance.

*firstname.lastname@aalto.fi

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

2. RELATED WORKS

The aspect of collaboration in the music experience and the creation of collaborative interfaces for music making has been extensively addressed earlier (see for instance [2]). Blaine [1] provides a framework for digital music instruments and systems which aim at supporting and taking advantage of the collaboration among players but also players and audience. In this respect, this work focuses on players' involvement, co-located collaboration, solely producing sound, for a group of three people or more.

Evaluating the design of a digital musical instrument is by no means an easy task. The difficulty mainly arises due to the complex nature of the experience of music making which can be hardly systematized as a set of tasks in which the activities involved can be measured quantitatively. Nevertheless, several authors are directing their sights towards the methods of human-computer interaction (HCI). In the last years HCI addressed the lack of paradigm able to fit the domain of non task-oriented computing but rather experience-oriented and a third paradigm of HCI focusing on embodied interaction, meaning and meaning construction, in specific contexts and situations is believed to be emerged [5].

More specifically for the field of new instrument for musical expression, O'Modhrain proposes a framework for the evaluation of digital music instruments [8]. Depending on the perspective on the design, different stakeholders, namely audience, music-maker (performer/composer) and designer, differently evaluate and shape the final design. Wanderley and Orio [12] proposes a qualitative approach for the evaluation of interactive musical system that focuses on a series of musical task of maximal simplicity users need to take. The tests are followed by a self-assessment of performance expressed on a Likert-scale. Kiefer et al. [7] investigated the Wanderley and Orio approach, and find qualitative analysis of interview data to be more useful than quantitative data about task accuracy.

Among the research interested in the users' conceptualization of musical interfaces, Stowell [9] proposes an evaluation method based on the tradition of Discourse Analysis (DA) and adapted to the experimental context. Also based on qualitative methods of analysis, Johnston presents an approach to practice-based research in new musical instrument design [6]. Our approach follows Johnston's recommendations and borrows most of his approach, namely the group-session, the semi-structured interview and the data-

analysis. The approach proposed here is a mix of qualitative data gathering and a modified think-aloud method.

3. AUGMENTING SOCIAL INTERACTION IN MOBILE MUSIC-MAKING

A system can be created which emphasizes the gesture-driven, ancillary communication between players while allowing their instruments to evolve according to group dynamics, spatial and temporal relation of behavior among the players. Such a system would emphasise eyes-free and gesture-driven interaction with instruments and take group dynamics into account. The extra-layer of sonic interaction should encourage human-human bodily interaction without distracting or inhibiting musical intentions of any of the promoters. The research aims at providing strategies for such an augmentation and methods for their evaluations.

3.1 Design strategy

The music controller designed is a mix of gestural controller and sonification of movement. Each player generates the sound autonomously from the other players, using the mobile phone and its body as an instrument. Parameters such as acceleration of the hand in 3-axis, quantity of motion of the whole body, contraction index and so forth, can be mapped in complex manner to a sound synthesis module.

Together with parameters that calculate the individual bodily engagement in the musical experience, we are interested in relational parameters that provide an indication of the group dynamics and social interaction while playing together. The relational descriptors are derived by combining together the spatial descriptors of each participant. Among those, the simplest and immediately useful is the distance among the players. As a result, the process obtained is authored and controlled as a group. Each player on one hand can play their own instrument but also needs to negotiate through bodily interaction and spatial disposition the overall sonic qualities of the music making.

3.2 Implementation

The three instruments were designed in order to make particularly explicit the gesture-sound relation. They were all percussive to keep the ensemble as symmetric as possible though different timbres were chosen spanning from bell-like sounds towards more wooden hollow box. For the mapping design, we followed principles drawn by [11], with respect to the tri-partite model consisting of binary mapping, basic parametric mapping and expressive mapping. For the latter two the magnitude of the accelerometers was used in the gesture-sound mapping while the touch screen input was used for the binary mapping. The gesture to sound link was created by a many-to-many mapping providing continuos control of the percussive sound. The sound synthesis was programmed using the RJDJ software[1] on iPhone.

We decided to position the speaker on the chest of the participant. This design was motivated by the fact we wanted to remind the user of the importance of using bodily interaction and movement in space in the improvisation and not to focus solely on the device.

For what concerns the augmentation, the system combines the use of different sensing technologies and its implementation is still in-progress. A Java software can receive the different skeletons tracking data from a motion capture system or a Kinect sensor and other signals sent via *open sound control* (OSC) from other devices. By tracking the horizontal position of the participants in the space, the system calculates the perimeter of the triangle formed by the participants. The perimeter relational descriptor is sent to the mobile phones and inversely mapped to the amplitude of a distortion filter applied to the instrument. As a result when the participants are far from each the sound of the instruments would not be affected by the relational descriptors but the closer they would get, thus reducing their perimeter, the higher the effect of the distortion could be heard. When two players are close but one is further away, the result is a modest distortion due to the fact that the position of the far away person has control over two sides of the triangle.

[1] http://rjdj.me/

4. USER STUDY AND EVALUATION APPROACH

The study involved N=21 participants, unpaid volunteers, 8 being female and 13 male, aged between 25-48, later divided in 7 groups of three players at the time. Musical background among the participants was even, varying from none to more than 10 years of playing several instruments. 15 of the participants previously played in bands and 11 declared at least some experience in electronic music making. 12 had previously used smart-phones as musical instrument.

In scenario 1 the sound of the three instruments was not affected by the distance parameter. In scenario 2, the influence of the distance among the players was switched on. Participants familiarized themselves with the instrument for 5 minutes. Afterwards, the user test started (Figure 1).

In both scenarios participants are presented with a complex, multimodal and unfamiliar situation, asked to do something new and together with people they do not know. Being in such a unnaturalistic setting further motivates us to investigate the experience as a whole rather than analyzing it along predetermined dimensions. Therefore we applied a mix of evaluation methods including quantitative survey analysis and qualitative interview data analysis.

4.1 Survey

The survey is here used as a light subjective evaluation regarding playability issues and aesthetic aspects of the sound that the instrument can produce. Moreover, in order to understand the main aspects emerging from the experience, we asked the participants to select the 5 most representative adjectives from a group of 20 chosen by us as a mix 10 positive and 10 negative adjectives: enjoyable, social, interactive, frustrating, engaging, delightful, awkward, energetic, disappointing, satisfying, active, organized, chaotic, pleasant, horrible, approachable, intimidating, welcoming, friendly and hostile. Participants positively addressed the overall experience rating **interactive** as the most relevant descriptor (14/21), followed by **active** and **social** (13/21).

We deliberately chose not to randomize the order of the two scenarios presented to the users. The reason behind this decision is that Scenario 2 can be seen as Scenario 1 plus the extra distortion parameter controlled by the perimeter. We feared presenting Scenario 2 as first being overwhelming because of the complete unfamiliarity with the setting and the instrument. As a result, a learning effect should be taken into consideration. We will explore this theme as connected with the one of familiarity in the analysis.

For the sake of comparison among the two scenario, we also asked an overall rating of the experience in the two scenarios with an adjective-anchored Likert-scale. Though participants preferred scenario 2 better than scenario 1, a Wilcoxon matched pairs test indicates a failure to reject the null hypothesis at the 5 percent significance level (p=0.146). This result obtained is not particularly surprising for a small pool of participants and in any way would not tell us much

about the experience of the user. For this, we need to proceed with an extensive interview as described in our evaluation approach.

4.2 Group interview and data analysis

The goal of the interview is to understand how users conceptually integrate the system into the context of use. To overcome the problem of assessing the players experience after the performance session, we adopted a Modified Stimulated Retrospective Think-Aloud (MSRTA), as suggested by [10]. The method is based on recording the user experience with video cameras and playing it back to them afterward in order to refresh their memory and stimulate a more grounded assessment of the session.

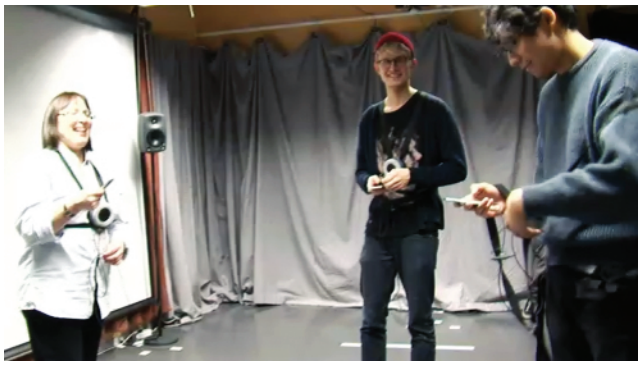

Figure 1: A still from the footage recorded in one user test session

We used techniques of data analysis from the grounded theory tradition [3]. The first part of the procedures is to code all the footage available, identifying and labeling incidents in the data, the so-called open coding phase. As a result of the coding for different incidents and actions, and elimination of obvious redundancy, the analyst reaches a smaller number of concepts. In the next step of axial-coding, the analyst searches for relationship among categories and sub-categories and through continuos comparison of incidents in the data, the process tends to converge towards the building of a theory.

Data consisting of roughly 160 minutes of interviews was coded and analyzed with the software Atlas.ti[2]. Memos for the video data annotated by the four authors independently were also used as primary documents. From an initial set of 113 codes grounded in 128 quotes, by drawing relationship among them, we achieved a final coding structure of highly dense[3] categories. From there, we sorted and extracted the core-categories: familiarity with the instrument and the situation, family of interaction, focus targets in the activity and categories of the experience.

Familiarity, associated with knowing how the instrument worked and had the time to learn it, was a pre-requisite to reaching the necessary control of the instrument and being able to play with others. People felt they needed more time to "experiment with the instrument" and even participants with extensive musical background felt "unusual" to play together with strangers and instruments they did not know in advance. For this, participants felt more comfortable with the situation in Session 2 due to familiarization with the instrument, the setting and the group happened in Session 1. Not randomizing the order helped easing this process and make the difference among the scenarios more explicit and understandable to them. Issues of controllability, internalization of the gesture-sound relation and expressivity also arouse.

In both scenarios placing the speaker to the chest was described both as "physical" but also problematic for the scope of hearing oneself instrument.

For what concerns "family of interaction" , 4 main categories were spotted: *interaction with others*, *with the instrument*, *with the device* and *among the devices*. The fact the interviewees refer to the same artifact with both the name "device" and "instrument" is quite significant. Under the core-category of "experience", the categories emerging are presented in Figure 2 that shows that one of the concept we encountered is the missed-link and, for some participants, the dichotomy between music-making and game. This result in relation to the other core-category "shift of focus" are the most relevant and will be elaborated further in the Discussion Section.

Figure 2: Categories of experience

5. DISCUSSION

We discuss below the other two core-categories and report some of the data grounding the interview analysis.

Family of interactions and Shift of Focus reflect participants' comments about either concentrating on the device or other players, often (though we can not say tendentially) in an exclusive manner.

"It is so easy but I am used to look at the device" (ID5)

"The device didn't interact" (ID6)

The ownership of the instrument and the centering effect of the device tends to work against the controlling-as-a-group strategy we proposed.

"I was thinking where to move and not listening" (ID20)

We believe this is one further reason, other than familiarity with the instrument, why movement shifts the focus away from a music-making activity. To counter that, one could focus on the interaction among the devices so to say mapping the distance among the devices rather than between the bodies.

"You focus too much on the distance" "you make the music with your hand but did not extend to the body" (ID12)

Nevertheless, our approach is different and wants the augmentation to stimulate participants to focus on each other rather than let the device guide them. For that, we believe movement should affect sound differently from what the player can achieve individually with the instrument. In this way, the augmentation becomes an affordance of the system that can be unveiled or ignored depending on the musical agenda of the individual player. Moreover, sonification of bodily relational descriptors without the use of mobile phones and the context of dancing might be alternative cases to explore further.

"You were concentrated on what happens when you interact with each other" (ID7)

'The social aspect...before it was difficult to look far from the device" (ID3)

In **categories of experience**, which encompasses issues connected with the overall experience, group-controlling seems associated with a game experience.

[2]http://www.atlasti.com/

[3]The concept of density is associated to concepts which are often grounded and highly connected to others.

"It was fun, I lost sense of time...I felt like a kid...playing with the movement...playful" (ID13)

Figure 3 shows our interpretation of the processes observed in the two scenarios as opposed to an ideal case of musical group improvisation. Familiarity with the instrument in the sense of being able to control it and modify its sound parameters is a necessary condition to manage the focus among the activity of playing the instrument and contributing and collaborating in the music-making process. As predictable, in Scenario 1, due to the unfamiliarity with the instruments, participants tried rhythmic patterns because simple musical ideas they would easily manage without losing focus on each other. In Scenario 2, while the unfamiliarity with the instrument partially persisted, the focus shifted away from the making-music to the curiosity of finding the cause-effect relation distance-sound and participate in a game-like activity. Our interpretation is that the two interaction paradigms for making sound, instrument and movement, were difficult to manage at the same time. The focus on listening to the others dropped in favor of an attention for each other movement and position in space, where the visual focus dominated the auditory focus, in turns not providing an easy cue of musical intention.

On the other hand, while departing from the initial music experience, the bond established while moving in space, being physically engaged and playfully blurring the edge between personal and social space led the participants feel a stronger social interaction and a sense of accomplishment from the standpoint of group activity (bottom of Figure 3).

"3 persons become one big instrument: interesting but restricting" (ID15)

Though more studies are needed, one might say making-music quits being the main goal of the activity but rather supports and posit the rules for a social interaction made of bodily and situated engagement. This finding motivates us retargeting our design strategy of augmentation for novices and casual players. For the issue of familiarity with the instrument, longitudinal study should be adopted by giving the non-augmented instrument to the players beforehand.

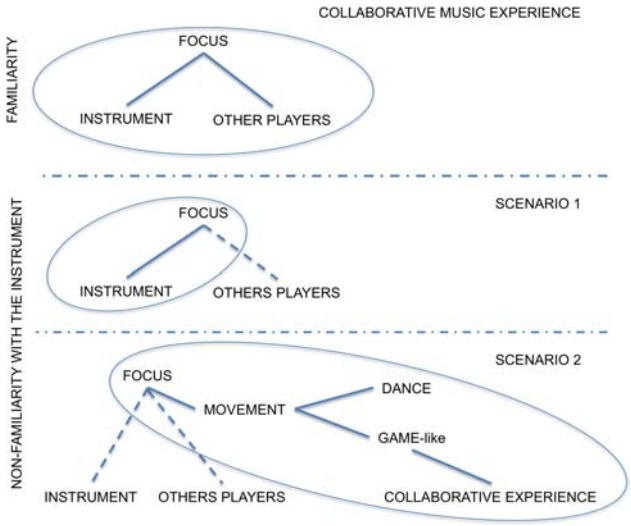

Figure 3: Relationship between shift of focus in activity and conceptualization of the experience

6. CONCLUSIONS AND FUTURE WORKS

In this paper we showed strategies of how to emphasize the social interaction naturally happening in collective music improvisation. The context of casual music making with mobile phone instruments was targeted. Our design approach take advantage of interaction cues such as distance among the players and relational descriptors to create musical affordances. An evaluation method was proposed that integrates modified think-aloud data gathering and analysis based on the Grounded Theory approach. The qualitative data analysis shows the suitability of our approach for augmenting the social dimension of music making enriching the playful interaction aspect of the group activity. Moreover, we underline the usefulness of the methodology for the evaluation of the experience and suggest both the augmentation strategy and the evaluation to be suitable for other context of collaborative music making.

7. ACKNOWLEDGMENTS

This work is partially supported by Aalto University Media Factory and Academy of Finland project 137646.

8. REFERENCES

[1] T. Blaine and S. Fels. Contexts of collaborative musical experiences. In *Proceedings of the International conference on New interfaces for musical expression*, pages 129–134. Singapore, 2003.

[2] N. Bryan-Kinns and P. Healey. Decay in collaborative music making. In *Proceedings of the International conference on New interfaces for musical expression*, pages 114–117, Paris, France, 2006.

[3] J. Corbin. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13, 1990.

[4] G. Essl and M. Rohs. Interactivity for Mobile Music-Making. *Organised Sound*, 14(02):197, June 2009.

[5] S. Harrison, D. Tatar, and P. Sengers. The three paradigms of HCI. In *Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA*, pages 1–18, 2007.

[6] A. Johnston. Beyond Evaluation : Linking Practice and Theory in New Musical Interface Design. In *Proceedings of the International conference on New interfaces for musical expression*, pages 280–283, Oslo, Norway, 2011.

[7] C. Kiefer, N. Collins, and G. Fitzpatrick. HCI methodology for evaluating musical controllers: A case study. In *Proceedings of the International conference on New interfaces for musical expression*, Genova, Italy, 2008.

[8] S. O'Modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):28–42, 2011.

[9] D. Stowell, A. Robertson, N. Bryan-Kinns, and M. D. Plumbley. Evaluation of live human-computer music-making: Quantitative and qualitative approaches. *International Journal of Human-Computer Studies*, 67(11):960–975, 2009.

[10] P. Strain, A. Shaikh, and R. Boardman. Thinking but not seeing: think-aloud for non-sighted users. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 1851–1856. ACM, 2007.

[11] A. Tanaka. Mapping out instruments, affordances, and mobiles. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 88–93, Sydney, Australia, 2010.

[12] M. M. Wanderley and N. Orio. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, 26(3):62–76, Sept. 2002.