Interactive reflexive and embodied exploration of sound qualities with BeSound

Giovanna Varni
DIBRIS-University of Genova
Viale Causa 13
16145 Genova, Italy
giovanna@infomus.org

Gualtiero Volpe
DIBRIS-University of Genova
Viale Causa 13
16145 Genova, Italy
gualtiero@infomus.org

Roberto Sagoleo
DIBRIS-University of Genova
Viale Causa 13
16145 Genova, Italy
sax@infomus.org

Maurizio Mancini
DIBRIS-University of Genova
Viale Causa 13
16145 Genova, Italy
maurizio@infomus.org

Giacomo Lepri
DIBRIS-University of Genova
Viale Causa 13
16145 Genova, Italy
lepri.giacomo@libero.it

ABSTRACT

The embodied and reflexive interaction paradigms separately proved to be effective for learning music in childhood. However, nowadays, there is a scarcity of research addressing the joined adoption of these paradigms, both from a theoretical and a technological point of view. BeSound supports children to explore - by means of their own body - rhythm, melody, and harmony and to creatively combine them together. Firstly, the child is engaged in a game in which she has to mimic the movement of the characters of stories BeSound tells her; then she can ad-lib a music dialogue with the characters. Each character was previously associated with a component of Laban's Effort and it was described through a set of whole-body movement features. These features are automatically detected, analysed, and used to control the music response of BeSound.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Representation**]: User interfaces—*Interaction styles, User-centred design, Auditory feedback*; K.3.1 [**Computers and education**]: Computer uses in education—*Computer assisted instruction*

General Terms

Human factors, design

Keywords

Reflexive interaction, embodied interaction, music education

IDC'13 Jun 24-27 2013, New York, NY, USA
ACM 978-1-4503-1918-8/13/06.

1. INTRODUCTION

Design of technology for learning, and in particular, for music education has been increasingly and effectively influenced by the *embodied* and *reflexive interaction* paradigms. O'Malley and Fraser [6] presented a literature review about the benefits of embodiment and physical activities in learning of young children. They highlighted that the use of physical *metaphors* can improve the learning rate of children. More specifically to music education, Antle and colleagues conducted a comparative study showing that embodied interactive environments could improve abstract music concepts learning in children [3]. They compared two environments: in the first one, body movements are mapped to music using embodied metaphors, e.g., an *up* movement is associated with *higher* volume; in the second one, a non-metaphoric mapping is performed, e.g., *more* activity is associated with *higher* volume. In the first environment, children demonstrated a higher performance level in conducting several tasks, e.g., vary the music volume from loud to quiet. A further example is *BodyBeats* a suite of three prototypes aimed at recognising, mimicking, and creating patterns through whole-body activity [9].

The reflexive interaction paradigm grounds on the idea of letting users mould virtual copies of themselves, through specifically designed machine-learning techniques. In the music domain, systems adopting such a paradigm are called *Interactive Reflexive Musical Systems (IRMS)*. They mainly focus on the *difference* between what the child can produce *with* the system and what the child can produce *without* it, rather than on the *quality* of what is produced [7]. Reflexive interaction and IRMS have been used in music education with children [1]. The first IRMS was *The Continuator*, a tool able to interactively learn and reproduce music of the same style as a human playing a keyboard [7]. At the present, the EU-ICT Project MIROR (Musical Interaction Relying On Reflexion, `www.mirorproject.eu`) [2] investigates the possibility of developing innovative adaptive systems for music learning and teaching in childhood, based on the reflexive interaction paradigm. In such a framework, *BeSound* was conceived and it is preliminary described in [8]. In particular, BeSound is part of *MIROR Body-Gesture*, which is a component of the *MIROR Platform*. The plat-

Figure 1: A frame of the animated movie in which the wizard engages the child in the game. Graphics by Compedia.

form includes two further components, *MIROR Impro* and *MIROR Compo*, devoted to improvisation and composition respectively. BeSound is aimed at supporting children in music learning, by exploiting storytelling to enable a cross-modal dialogue with the machine, which is both embodied and reflexive. More specifically, BeSound supports children to explore - by means of their own body - rhythm, melody, and harmony and to creatively combine them together. Firstly, the child is engaged in a game in which she has to mimic the movement of the characters of stories the system tells her; then she can ad-lib a music dialogue with the characters. Each character was previously associated with a component of Laban's Effort and it was described through a set of whole-body movement features. These features are automatically detected, analysed, and used to control the music response of BeSound. A description of Laban's Theory of Effort can be found in [4] and [5]. The *Space*, *Time*, and *Weight* components of Effort were associated with the basic components of music composition: melody with Space, rhythm with Time, and harmony with Weight, respectively. Due to the relatively scarce literature on this specific topic, the proposed association emerged from discussion between experts of Laban Movement Analysis, professional musicians, and psychopedagogists.

The remainder of this paper is organised as follows: Section 2 illustrates the narrative context the child is provided with, Section 3 describes the overall architecture of BeSound, Section 4 provides the some technical details of how it works.

2. BESOUND NARRATIVE CONTEXT

A short animated movie, set in a magic world, was chosen to provide the child with a concrete and friendly context and to motivate her to play. In this movie, a wizard introduces himself and tells that his students, by mistake, made the land of make-believe loose all of its sounds. Figure 1 shows a frame of the movie.

Then, he asks the child to help him to give the characters their sound back: he will make a magic powder turning movement into sound, so he needs that the child mimics the movement of the character acting in its own story. At this point, the child can select the characters she would like to help. Once the selection is completed, the stories with the chosen characters are told one after another and the child tries to mimic the movement of the characters. At the end, the wizard congratulates the child and, to celebrate the recovery of the sounds, organises a party. However, there are not musicians at all. The wizard, with the help of the child, can make a new magic: the child will mimic again the movement of the characters and their sounds are transformed into the sounds of musical instruments. The child can now play and engage in a musical dialogue with each of characters.

Mimicking the stories, acting to transform the sounds of the characters into the sounds of musical instruments, and finally having a dialogue with the characters correspond to different working modes of BeSound. Modes are conceived to reach pedagogical objectives: the first one is devoted to encourage children to listen to the sounds and to focus on them and on the relationship between movement qualities and sound; the second one is aimed at introducing different musical instruments and at learning that musical instruments have different timbres; finally, the third one is intended to make children interact in real-time with sound in a kind of guided improvisation, and to develop the ability of working with sound material with critical sense. Mimicking is also devoted to build a model of the Effort of the child. All the graphics was provided by Compedia `http://www.compediamoodle.com/education/`.

3. OVERALL ARCHITECTURE

BeSound consists of the following components: teacher interface, children interface, kernel, storytelling-dialogue manager. It uses a Kinect sensor as input device.

3.1 Teacher interface

The teacher interface allows the teacher to configure and handle a BeSound session. This interface was developed in Java and communicates both with a remote database storing particulars of the children (e.g., name, surname, and information on whether a child already played with BeSound) and with the storytelling manager. The interface also allows the teacher to choose the language in which the stories will be told (English, Italian and Swedish versions are available) and the BeSound working mode to be administered to the child. Finally, the selection of which and of how many characters the child can select and which stories she will act is enabled. Once the teacher confirms all her choices, these are sent via network to the storytelling manager, which invokes the children interface. The teacher can decide to quit the current BeSound session whenever she likes.

3.2 Children interface

The children interface, customisable for each child and for each session, is designed as a storybook where the characters are placed in. Figure 2 shows a snapshot of it.

The live image of the child is captured and pre-processed by the Kinect sensor. It is then projected inside the book. When the child is detected standing in the middle of the picture of the storybook, she can select her favourite characters. The detection of the child position is done by comparing the current coordinates of the torso of the child, received from the Kinect sensor, with the coordinates of the centre of the book: the 2D projection of the torso of the child has to belong to a region around the centre of the book. Selection of characters is performed with a free movement of the child's hands, i.e., the child can select a character by moving her hand on the projected character. Even if the touch is not real, the child can see her projection (her mirrored image)

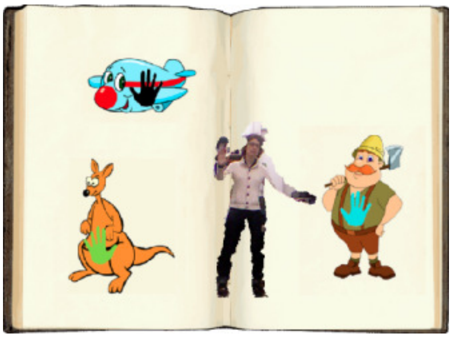

Figure 2: Children interface: the child chooses by her hands her favourite characters in the storybook. Graphics of characters by Compedia.

that touches the character on the screen. If the contact lasts for 3 seconds at least, the touch is recognised as valid. The chosen character disappears in order to avoid a possible double selection of the same character. The 3D coordinates of the child's hands are mapped onto the 2D display of the children interface and the 2D distance between the position of each hand and the position of each character is measured. To allow for some inaccuracy in the selection action, a region around the centre of each character is considered as valid for selection. A coloured hand shape is drawn over every character to help the choice. Once the child performs the number of selections the teacher established, the children interface disappears and the first story starts. The numeric *ids* of the chosen characters are sent via network to the storytelling manager only when the selection is completed.

3.3 Kernel

The BeSound kernel includes different software modules. A *movement detection and feature computation module* extracts the representative features of the movement of each character from the MoCap data the Kinect sensor captures. A *movement analysis module* implements a Kohonen's self-organising map for classifying movement in terms of Laban's Effort components. Finally, a *dialogue engine* manages storytelling and the dialogue with the child. The kernel is implemented via software by using EyesWeb XMI. Due to the different movement features chosen for each character, a specific BeSound kernel instance was developed for each of them. Each instance communicates with the dialogue engine by using the interprocess communication facilities the RCF (Remote Call Framework) framework provides.

3.4 Storytelling-dialogue manager

The storytelling-dialogue manager is aimed at managing: (i) the development of the overall narrative context, (ii) the flow of the stories of the characters; and (iii) the dialogue with the characters. It is implemented as a tuple of interconnected finite-states-machines able to send/receive data to/from the kernel. More in details, there is one supervisor finite-states-machine and one finite-states-machine for each character. All the finite-states-machines were developed in Python and use the MetaEyesWeb platform. The supervisor manages the activation of and the communication with the teacher interface and the children interface. Moreover, it manages the activation of the finite-states-machines handling each single character. All the finite-states-machines handling the single characters carry out the following operations: they manage the flow through the stories as described in the next section, and the turn-taking of the reflexive music dialogue with the characters.

4. WORKING MODES

This section summarises how BeSound works following the narrative context presented in Section 2.

4.1 Mimicking characters

Every story involves a character at a time. At the present, the following three characters, one for each Effort component, are available: the woodman, the kangaroo, and the airplane addressing the Time, Weight, and Space Effort component, respectively. Every story consists of four paragraphs. The first paragraph introduces the character and the context she acts in. The second and the third paragraphs describe the actions the character performs in her story. Finally, the fourth paragraph gives a conclusion. The child is asked to mimic the actions in the second and third paragraphs only. At the end of every of these two paragraphs, the child either moves to the next paragraph or repeats the previous one depending on the result of the automatic features analysis and Laban's Effort component classification (i.e., whether the character's actions were mimicked with the correct Effort qualities or not). Further, when the current paragraph ends, the child can listen to the sound associated with the actions the character performs. This sound depends on the Effort component (that is, on the character the child is mimicking), on its quality and its magnitude. Each Effort component has two qualities (e.g., *quick* and *sustained* for Time component) and for each of these two levels of magnitude are taken into account. In more detail, the movement features are computed by using a sliding-window approach in which the size and the step of the window depends on the character. Then, they are provided as input to a Kohonen's self-organising map previously trained with the same movement features extracted from a corpus of videos involving six children acting the same stories told by researchers.

At the end of each session, the teacher can decide to update the Kohonen's self-organising map with the features of the child: in such a way BeSound builds a custom Effort's model for each child. This model is exploited in order to implement the reflexive music dialogue.

4.2 Transforming sounds

This mode is aimed at leading the child to the reflexive music dialogue by presenting her musical instruments in place of characters' sounds. The wizard asks the child to choose again the characters she would like to play with and to perform again the movements she learnt acting the stories: the computed features are the input of the Kohonen's self-organising map, customised for that specific child. The classification of the actions the child performs in the first three seconds decrees the resulting sound the child listens to. In this case too, there are four different sounds for each Effort's component. The chosen musical instruments are a saxophone, a cello and a marimba. They are played by the airplane, the woodman, and the kangaroo, respectively. Instruments were selected taking into account similarities with the sounds of the characters with respect, e.g., to timbre and attack. For example, plucked cello strings are characterised by a sharp attack recalling the woodman's axe hitting a tree, whereas the bowing sound of the cello resembles the slower and long sound of the saw.

4.3 Reflexive music dialogue

The wizard encourages the child to engage a music dialogue with the characters. He proposes to start by imitating the music they play. The storytelling-dialogue manager processes the Kohonen's self-organising map of the child and generates a music sequence according to the Effort's quality distribution over the classes. More specifically:

- The *length* of the sequence depends on the extent to which the motoric activity of the child while mimicking the stories was high. The more the child performed actions the more the sequence is long. The minimum length of sequence is seven short music fragments;
- The *music fragments* composing the sequence are chosen to flat as much as possible the Effort's quality distribution resulting from the Kohonen's self-organising map, i.e., to induce the child to explore all the available sounds and actions in an uniform way. The first stimulus belongs to the most acted Effort's quality in the distribution, the other ones are almost inversely proportional to the extent to which each Effort's quality is acted, that is, the Effort qualities and the related sounds that the child rarely explored will appear in more items of the sequence.

The turn-taking of the dialogue is managed as follows: the character proposes the first stimulus of the sequence and, then, the child tries to imitates it by acting as long as she wants the character's movements. The customised Kohonen's self-organising map classifies the currently performed Effort quality, which is mapped onto the audio output. Thus, the child can listen to the sound produced by her own actions while she is performing them. The sounds are those of the musical instruments previously introduced. When the child stops, the most prevalent Effort quality she performed is computed. The response of the character is then produced so that it starts by repeating the most prevalent quality and then varies to reach either the next item in the sequence or the current item, in case the child did not imitate the character. Such a mechanism of repetition and variation is one of the pillars of reflexive interaction [7].

During the dialogue the distribution of the classified Effort's quality is updated so that the generation of a possible new music sequence will take into account how the child proceeded in exploring Effort qualities and related sounds.

Further kinds of music dialogue are currently under investigation such as one enabling the child to start proposing a music hint and then play more freely with each character, possibly emphasising variations. Simultaneous playing with other children is also planned: up to three children will mimic the movements of three characters, associated to the three music components of melody, rhythm, and harmony. The sounds they produce will be combined in a single music composition.

5. CONCLUSIONS

This paper described BeSound, a crossmodal embodied reflexive application aimed at exploring rhythm, melody, and harmony by means of whole body movement. Reflexivity is particularly addressed in the music dialogue. Indeed, in such a dialogue the mirroring effect is obtained both in the crossmodal relationship between movement and gesture and in the content of the system's music responses which include repetition and variation. Moreover, BeSound automatically adapts to the child's style both with respect to the motoric activity (update of the Kohonen's self-organising map) and with respect to exploration of sounds (update of the distribution of Effort qualities) so that the more the child plays with BeSound the more BeSound becomes a mirror of the child. According to the reflexive interaction paradigm, such a continuous evolution of the system should make its behaviour change along time. In IRMS this proved to support prolonged interaction in the long term.

6. ACKNOWLEDGMENTS

This work was partially supported by the EU-ICT Project MIROR (Musical Interaction Relying On Reflexion, `www.mirorproject.eu`). The authors thank the teams of University of Bologna (Anna Rita Addessi, Filomena Anelli, Rosane Cardoso, Marina Maffioli, Fabio Regazzi), of Compedia (Shai Newman and Michal Skurnik), and of University of Gothenburg (Åsa Bergman, Bengt Olsson, Sarah Mercieca) for the useful feedback. Special thanks to Antonio Camurri, Corrado Canepa, Paolo Coletta, Simone Ghisio, Barbara Mazzarino, Michele Pizzi, and Luca Serra for their precious contributions to this work.

7. REFERENCES

[1] A. R. Addessi and F. Pachet. Experiments with a musical machine. musical style replication in 3/5 year old children. *British Journal of Music Education*, 22(1):21–46, 2005.

[2] A. R. Addessi and G. Volpe. The miror project. In *Towards Ubiquitous Learning. Proceedings of the Sixth European Conference on Technology Enhanced Learning*, pages 15–28, 2011.

[3] A. N. Antle, M. Droumeva, and G. Corness. Playing with the sound maker: do embodied metaphors help children learn? In *Proceedings of the 7th international conference on Interaction design and children*, IDC '08, pages 178–185, New York, NY, USA, 2008. ACM.

[4] R. Laban. *Modern Educational Dance*. Macdonald and Evans Ltd., 1963.

[5] R. Laban and F.-C. Lawrence. *Effort*. Macdonald and Evans Ltd., 1947.

[6] C. O'Malley, D. S. Fraser, et al. Literature review in learning with tangible technologies, 2004.

[7] F. Pachet. Enhancing individual creativity with interactive musical reflective systems. In I. Deliège and G. Wiggins, editors, *Musical Creativity: Multidisciplinary Research in Theory And Practice*. Psychology Press, 2006.

[8] G. Volpe, G. Varni, A. R. Addessi, and B. Mazzarino. Besound: embodied reflexion for music education in childhood. In *Proc. 11th Intl. Conference on Interaction Design and Children, IDC '12, Bremen, UNK, Germany, June 12 - 15, 2012*, pages 172–175, 2012.

[9] J. Zigelbaum, A. Millner, B. Desai, and H. Ishii. Bodybeats: whole-body, musical interfaces for children. In *CHI Extended Abstracts*, pages 1595–1600. Springer Verlag, 2006.