Conveying audience emotions through humanoid robot gestures to an orchestra during a live musical exhibition

GIARDINA, Marcello, TRAMONTE, Salvatore, GENTILE, Vito, VINANZI, Samuele, CHELLA, Antonio, SORCE, Salvatore and SORBELLO, Rosario

Conveying Audience Emotions through Humanoid Robot Gestures to an Orchestra During a Live Musical Exhibition

Marcello Giardina[1,2], Salvatore Tramonte[1,2], Vito Gentile[1,3],
Samuele Vinanzi[1,2,4], Antonio Chella[1,2], Salvatore Sorce[1,3], and
Rosario Sorbello[1,2]

[1] Dipartimento dell'Innovazione Industriale e Digitale (DIID)
Università degli Studi di Palermo
Viale delle Scienze, building 6 - 90128 Palermo, Italy
{firstname.lastname}@unipa.it
[2] Robotics Lab
[3] Ubiquitous Systems and Interfaces Group - http://usi.unipa.it
[4] Centre for Robotics and Neural Systems - Plymouth University
PL4 8AA Plymouth, United Kingdom
samuele.vinanzi@plymouth.ac.uk

Abstract. In the last twenty years, robotics have been applied in many
heterogeneous contexts. Among them, the use of humanoid robots during
musical concerts have been proposed and investigated by many authors.
In this paper, we propose a contribution in the area of robotics appli-
cation in music, consisting of a system for conveying audience emotions
during a live musical exhibition, by means of a humanoid robot. In par-
ticular, we provide all spectators with a mobile app, by means of which
they can select a specific color while listening to a piece of music (act).
Each color is mapped to an emotion, and the audience preferences are
then processed in order to select the next act to be played. This decision,
based on the overall emotion felt by the audience, is then communicated
by the robot through body gestures to the orchestra. Our first results
show that spectators enjoy such kind of interactive musical performance,
and are encouraging for further investigations.

Keywords: HRI, Musical Robotics, Humanoid Robotics

1 Introduction

In the last decade, we have witnessed a growing interest towards the use of
intelligent humanoid robots during common human activities, because of their
very "human" aspect and consequent human-like motion capabilities. This gives
them the ability to convey emotions during the interactions with humans, and
this is the main reason that boosted up the development of humanoid robots
and their consequent availability for common-life applications. There are a lot
of robots that are specifically designed to convey emotions, and their shape is
often fine-tuned even on the specific emotional field they are intended for.

For example, emotional humanoid robots are used to sustain natural empathic conversations with elderly people [20], to express emotion and personality [6], to aid people with cognitive disorders [1] and to support disabled people such as "ALS" patients [21]. Furthermore, gestures may be suitably used both by humans for the live control of humanoid robots [9], and by humanoid robots to communicate in a "human" way [10].

In the field of the musical performance, we propose that Human-humanoid interaction (*HHI*) allows for designing natural-like settings in which human users and a robotic agent coordinate behaviours and emotions [2] and communicate on the basis of shared phenomenal features in a common environment [19]. Our goal is to build a bond between the two entities giving to a robot the capability to exhibit affective-expressive movements as an emotional mediator of the feelings of the human audience during the musical performance of an orchestra. In this paper, we present a system for conveying audience emotions during a live musical show, using a humanoid robot as an orchestra leader. In particular, we provide all spectators with a mobile app, by means of which they can select a specific color (which corresponds to an emotion) during a live performance. Then, audience feelings are used by the robot to address the musicians to play a piece of music with a specific emotional meaning according with the musical composition. The robot directs the orchestra using preset body gestures.

The rest of the paper is arranged as follows. Section 2 presents a brief review of the state of the art in terms of emotion representation and the use of robotics in music; Section 3 includes a detailed overview of the whole system; Section 4 shows the first results of an evaluation study conducted during an actual deployment of our proposal; finally, Section 5 concludes the paper and describes future works.

2   Related Works

Before describing our system, in this Section we focus on two crucial aspects addressed in this paper: emotion representation and robotics applications in musical context. The first subsection is aimed at describing which frameworks have been studied in literature in order to quantitatively model emotions. The second one includes an overview of robot-based solutions in musical contexts.

2.1   Emotion Representation

Emotions can be described as the human response to adapt himself to the constant environmental stimuli deriving from one's circumstances, mood, or relationships with others. Emotions are the results of complex psychological and physiological systems activated by the perception of significant environmental events. Many researches have been conducted in order to investigate how to represent emotions and how it is possible recognize them, as shown in [11] or [15]. Among the many approaches illustrated in psychology, the one based on "emotional categories" is one of the most often used [7] in literature, having focus to modelling emotions in according to different and discrete emotion classes
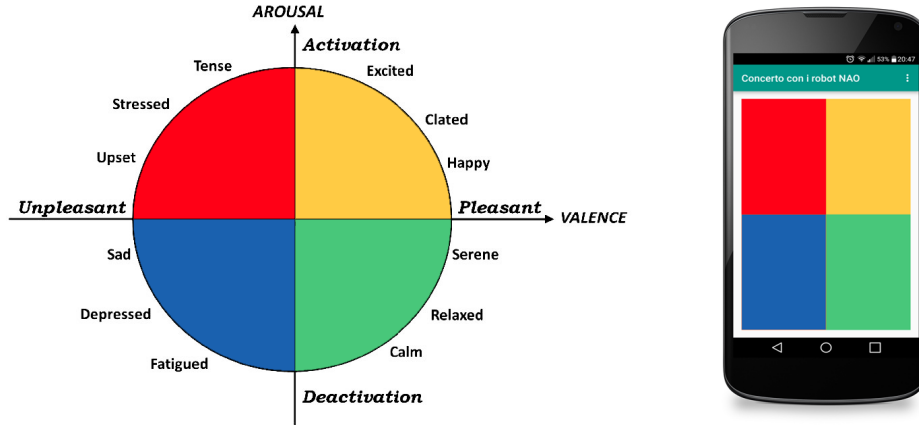
Fig. 1. The circumflex model for emotion representation, along with color labeling and its use in the mobile application.

and labels. One of the most used approaches has been developed and described by Paul Ekman in [8]. In his work, Ekman claimed that all the emotions can be mapped into a six-dimensional space defined by six basic emotions, namely *anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*. Using this model it is possible representing a human emotion felt as a balance of the six basic emotions. This advantage, however, may also represent the main drawback of such approach, since non-expert users may consider it overcomplicated to model their feelings. A two-dimensional, alternative to the Ekman's model, has been described by Russell et al. in [17], and the further extended by Posner et al. in [16]. In these works, authors defined a "circumflex model of affect", aimed to accurately describe every possible emotion. In particular, authors suggest a two-dimensional structure in which an emotional event may vary quantitatively along two independent dimensional variables. One dimension (*valence*) describes the degree of pleasantness or unpleasantness, whilst the other dimension (*arousal*) indicates the intensity in terms of physiological activation. The combination of these two dimensions generates the subjective emotional feeling. By means of this model, every emotion can be defined as the combination of two different levels of arousal and valence. The emotional categories may be ordered along the aforementioned circumflex, as shown in Figure 1.

In this work, we opted for the circumflex model to describe the audience emotional responses, as well as the emotional content of the pieces of music of the composition. This allows us for discretizing the whole spectrum of emotions in four "areas", one for each quadrant of the circumflex. Each of these areas was associated with a different color: from the first to the fourth quadrant, happiness (yellow), anger (red), sadness (blue) and serenity (green). These colors have been chosen for being associated with the correspondent emotions [3] [22].

2.2   Robotics in Music

There are several relevant works in the field of Robotics in musical applications. Among them, Hoffman et al. have defined the sense and outcomes of sharing an experience with a robot and this concept has been proposed as *Robotic Experience Companionship* (REC) [12]. Lim et al. proposed a unifying framework to generate emotions across voice, gesture, and music, by representing emotional states as a 4-parameter tuple of speed, intensity, regularity, and extent [13]. McCallum et al. stated that improvised musical interaction are able to provide an improvement of social presence and engagement during long term Human Robot Interaction (HRI) [14]. Burger et al. developed a simple robot with the goal to display its emotions by performing expressive movements during musical performances [5]. Brown et al. derived a framework for implementing happy and sad gestures on a humanoid robotic platform to measure user's predominant perceived emotions to enhances the social interaction [4]. This short list show the broad interest of the scientific community towards the use of (humanoid) robots to convey emotions both from and to people in a musical context.

3   System Description

In this Section, we describe the main building blocks of our system architecture, depicted in Figure 2.

There are three main actors that produce inputs or receive outputs: the *orchestra*, the *audience* and the *setlist*. The first one is made up of an electronic ensemble (i.e. a group of electronic acoustic speakers controlled by a computer) and an instrumental ensemble that produces music with acoustic instruments. The orchestra's role is to receive the system output in terms of gestures from the robot, and to play according to them.

The second actor of this system is the *audience*, i.e. spectators of the concert who are equipped with a mobile application (namely *NaoMusic*). As explained further in the following Sections, spectators are able to select their preferred colors according to the emotions felt during the musical performance.

The last actor is the *setlist*, i.e. a collection of audio files that together compose the musical structure of the concert. In particular, these musical tracks will be played by the electronic ensemble, acting as the base for the acoustic performance by the instrumental ensemble.

The system architecture is made up of five modules:

- *Emotion Recognizer*: a machine learning module to extract the emotional content from a song;
- *Emotional Mobile Interface*: a mobile app (i.e. the aforementioned *NaoMusic*) that allows users to submit their preferred color, according to the emotions experienced during the concert;
- *Database*: a web-located database that stores all the users' choices;
- *AI Director*: central component that realizes the decision-making process, guiding the evolution of the musical performance;
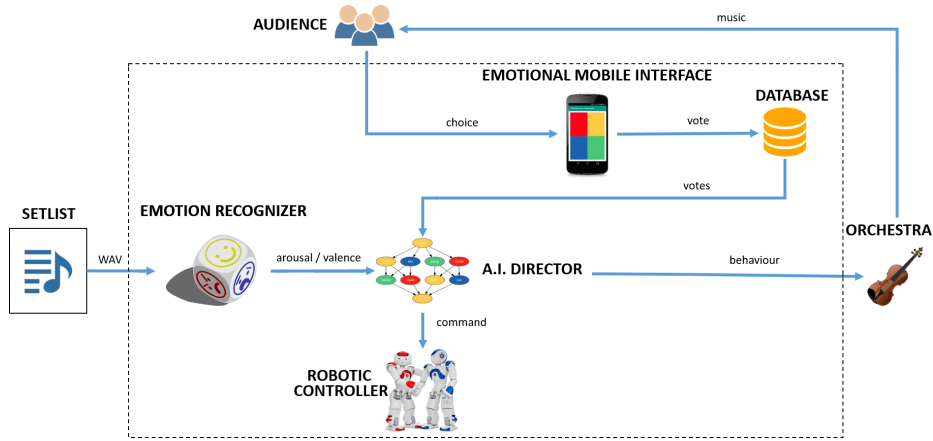
Fig. 2. System Architecture.

– *Robotic Controller*: translates the instructions of the *AI Director* module into behaviours for the robot.

Before describing the modules in more details, in the following subsection we present a high-level use case, in order to better clarify how the system works.

3.1 Use Case

The concert is preceded by a preparation phase aimed at generating the music to be played by the orchestra. In particular, some musicians compose a few short pieces, which is then processed by a software in order to produce a series of new, original tracks. All the new computer-generated songs, along with the initial ones, are included in the setlist used for the concert

All these tracks are thus sent to the Emotion Recognizer module, which extracts (as explained later) their emotional content in terms of arousal and valence, from here on called *song_arousal* and *song_valence*. The output of this functional block is thus a collection of ⟨*song_arousal*, *song_valence*⟩ pairs, one per track in the setlist. This concludes the preparation phase, and the system is now ready for being used in an actual concert.

During the concert, the orchestra starts by playing the introductory music, while the audience is able to express its emotions using the Emotional Mobile Interface (i.e. *NaoMusic*). All the data sent by this application are continuously stored in a remote Database. When the musical act is close to its conclusion, the AI Director queries the Database to fetch the entries gathered during the current act. Those data are then processed in order to estimate the overall emotional state of the spectators, in terms of a ⟨*song_arousal*, *song_valence*⟩ pair. By using this information, along with the emotional content of the tracks in the setlist, the system selects the song whose emotional content best fits the course

of the performance, communicating its choice to the Robotic Controller. This module translates the choice into a behaviour for the robot, which will inform the orchestra about the next song to be played.

In the following we describe each module in more details.

3.2   Emotion Recognizer
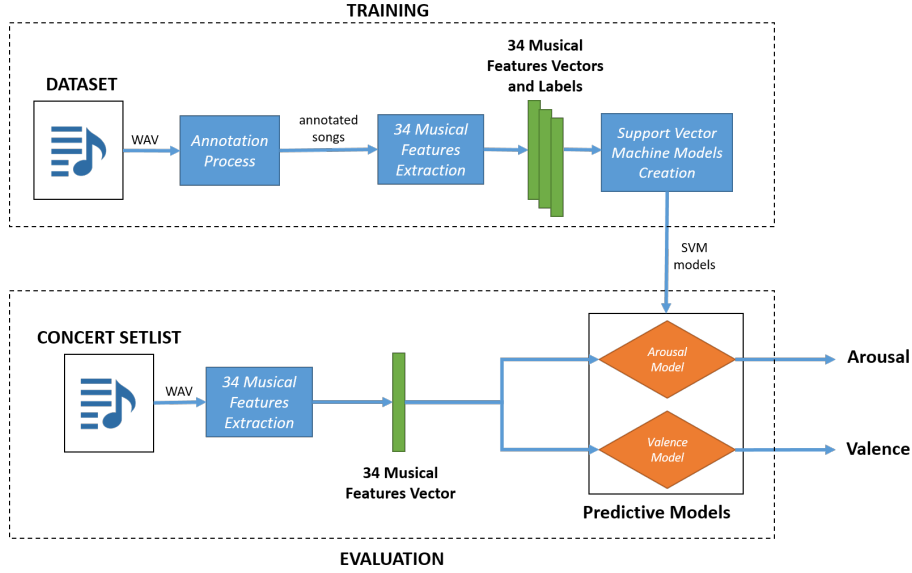
Fig. 3. Emotion Recognizer module architecture.

This module implements a method for allowing identification and extraction of ⟨*arousal*, *valence*⟩ pairs from a musical track. The emotion recognition capabilities have been implemented through Support Vector Machines (SVMs), trained with the dataset described in [18]. It includes approximately 1000 CC-licensed songs that have been listened and subsequently annotated with their values of arousal and valence through crowdsourcing. The training process uses 34 musical features extracted from each audio track of the aforementioned dataset, which constitute the training set inputs (see [18] for more details). Obviously, the corresponding *arousal* and *valence* annotations are used as outputs. More specifically, two models are created: one for arousal recognition, and the other for valence recognition.

After the training stage, the setlist is processed to extract the aforementioned musical features from each track. These features are then inputted into the two SVMs to obtain the estimated ⟨*arousal*, *valence*⟩ pairs.

3.3 Emotional Mobile Interface

This module plays a fundamental role in the proposed system, as it represents the interaction point between the audience and the entire system. It consists in a simple and user-friendly mobile application, whose interface is shown in Figure 1.
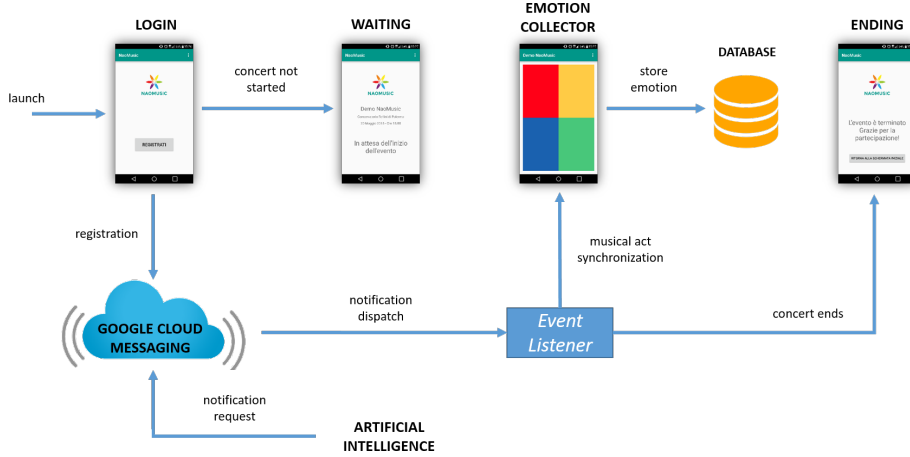

Fig. 4. Emotional Mobile Interface architecture.

Figure 4 describes the mobile application architecture. The main goal is to create an interface that allows spectators from the audience for expressing their emotional state while listening to the concert. All the votes recorded by all users at the end of each musical act will be used in order to compute the emotion that more faithfully represents the global emotional state of the audience, and that will be used by the AI Director module to decide which song should be played next.

On startup, the mobile devices can receive push notifications by the AI Director module through the Google Cloud Messaging (GCM) infrastructure. Once logged in, the application will move its user into a welcome screen that will persist until a notification will report the beginning of the concert. At this point, the user will be led in the voting screen, where s/he will be able to send the emotion to a remote webserver. This interface will synchronize with the musical act progression, thanks to the notification system, to be able to store the votes correctly. This screen resembles Russel's Circumplex Model of Emotions, having four colored tiles that resemble the four emotional categories described in Section 2.

It is worth noting that nor before neither during the concert users have been advised on the meaning of the colors: they were only asked to interact with the
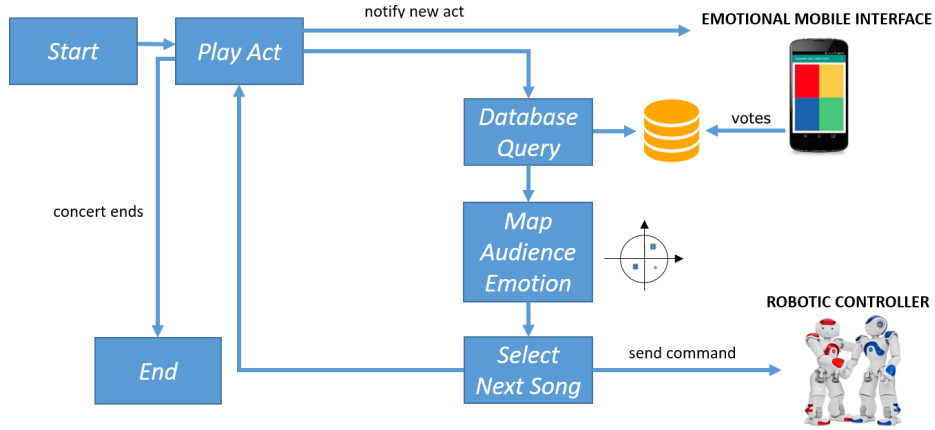
Fig. 5. AI Director architecture.

app as they wish, in order to avoid biases and to stimulate a more instinctive and less reasoned response.

3.4   Database

The Database module consists in a simple HTTP server, able to receive GET and POST requests aimed at storing votes to a local MySQL database management system. More precisely, all the votes are firstly sent from the mobile application to a server via GCM, and then stored to the database using REST API. The data are then used from the AI Director module (described in the next subsection), or for further offline analysis.
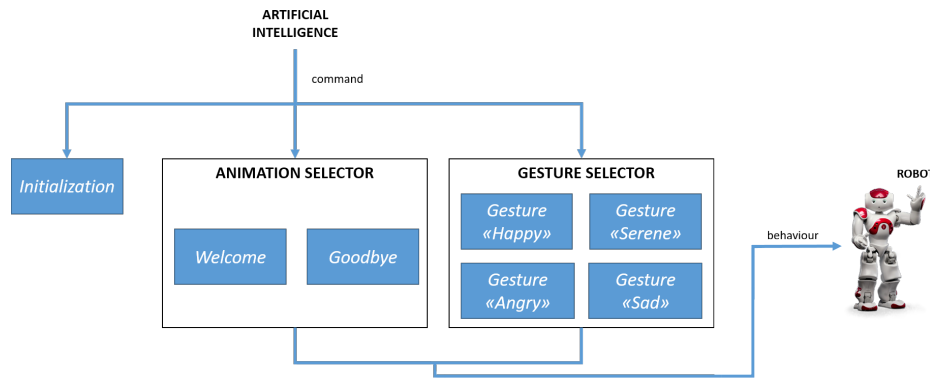
3.5   AI Director

This module is the core component, as it contains the artificial intelligence that deals with the real-time management of the concert, taking decisions on the next song to be played. This module accepts as input the ⟨*song_arousal*, *song_valence*⟩ pair produced by the Emotion Recognizer module.

In order to understand how this module works, it is important to explain the concert structure. It consists in a tree data structure, whose depth represents the number of musical acts, and whose nodes are all the possible songs that can be played in every act (see Figure 9). The goal of this module is to find an optimal path from the first to the last node, in order to produce the most engaging effect for the spectators based on their emotional response.

At the beginning of the show, the robot will perform an introductory welcome animation and subsequently it will turn towards the musicians in order to communicate via a body gesture the beginning of the first musical act.

For every such acts, the AI Director sends a synchronization notification to the Emotional Mobile interface, then it remains in a waiting state, giving time to the electronic and orchestral ensembles to perform their executions. Shortly before the end of the act, the system will wake up and query the Database to retrieve the spectators' votes collected during the act. This information is used to calculate the audience's predominant emotion. Combining this information with the emotional content of the candidate songs eligible to be performed on the next act, a heuristic search algorithm can be used in order to choose the song that most likely would induce an emotional shift in the audience, while avoiding stalls on the same emotions and abrupt changes. For instance, considering the $\langle song\_arousal, song\_valence \rangle$ pair as a point on the circumflex shown in Figure 1, its mirrored point (with respect to the main diagonal) represents the opposite emotion, which can be selected for obtaining the highest emotional variability.

3.6 Robotic Controller

Fig. 6. Robotic Controller architecture.

Once the AI Director chooses the next song (as explained before), the Robotic Controller module will convey this information to the robot, and through it to the performants. This chain of operations is executed in loop for every musical act, until the end of the show.

The purpose of this module, whose architecture is shown in Figure 6, is to model the robotic behaviors that will enable a NAO robot to interact through body gestures with people involved in the show, both spectators and performers.

The robot plays two main roles: it serves both as an announcer, gesticulating and speaking to the audience, and as a musical director, performing movements to instruct the musicians on the next song they should play. The gestures used for communicating emotions are shown in Figure 7, and have been designed to be non-ambiguous. These four movements are the robotic embodiments of the

outputs from the AI Director module, and each of them is interpreted as one of the four songs that can be played in the next musical act. Each song belongs to an emotional category, which in turn is associated to a color. Thus, the robot will modify the color of his LEDs accordingly, to give a continuous reference to the music players even after the gesture movement is over.
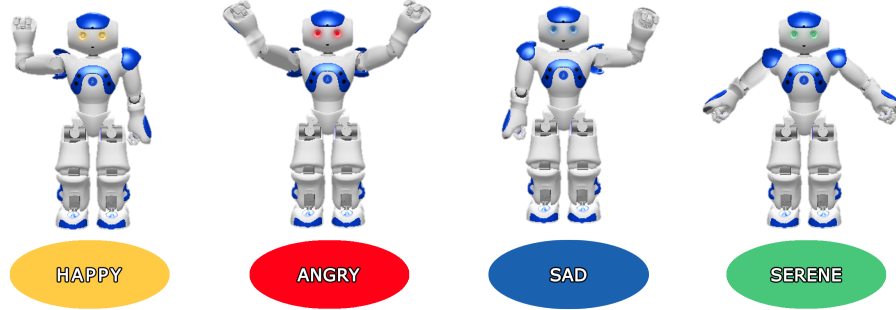

Fig. 7. The four robot directional gestures.

4 Evaluation

After some preliminary tests, aimed at fixing bugs and tuning some parameters, we tested our system during an actual concert, opened to an audience made by 118 spectators (63 males and 55 females). As explained before, each spectator was given with a mobile application, by means of which they were able to select a color during the concert. At the end of the concert, we have also asked them to fill a questionnaire consisting of 9 questions to be answered with 3-points Likert scales:

1. How frequently do you listen to symphonic music?
2. Do you consider yourself an expert in informatics?
3. Which was the predominant emotion you felt during the concert?
4. Did the robot correctly fulfil the task?
5. Do you believe that the presence of a person instead of the robot might change the effectiveness of the concert?
6. Before the concert, did you expect an emotional involvement from the performance?
7. Did the robot meet your expectations?
8. Did the presence of the robot affect your emotional state?
9. Did you enjoy the concert?
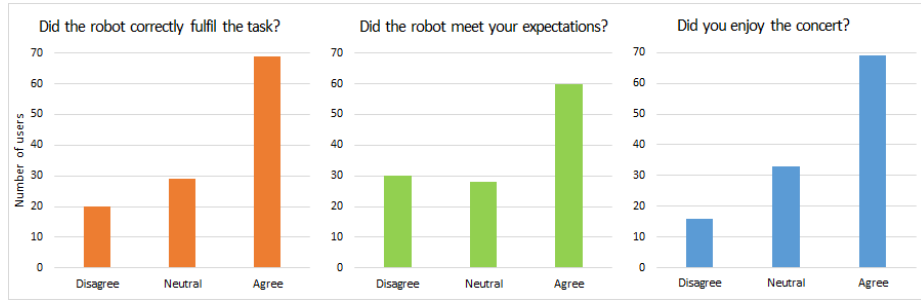
Fig. 8. Users' answers to questions 4, 7 and 9.

Most of the questions produced answers with mean and median values next to the neutral one. However, it is worth noting that questions 4, 7 and 9 revealed interesting aspects of users' opinions. As shown in Figure 8, the majority of users enjoyed the concert, and no relations were noted between neither the level of expertise in informatics nor the users' habits in attending to such kind of concerts. Moreover, most of the users felt like the robot correctly fulfil the task, and that its behavior meets users' expectations.

During the concert, the predominant emotion recorded was Serenity (i.e. most of the users chose green). Furthermore, the analysis of users' responses in terms of the chosen colors (see Figure 9) shows that they do not always select the emotions corresponding to the songs played by the orchestra. Indeed, according to the output of the emotion recognizer module, the songs played by the orchestra should elicit a specific emotion. The observed differences in users' choices may be due to external factors, including the presence of the robot. While this consideration needs to be further investigated, this preliminary result is encouraging for future works.
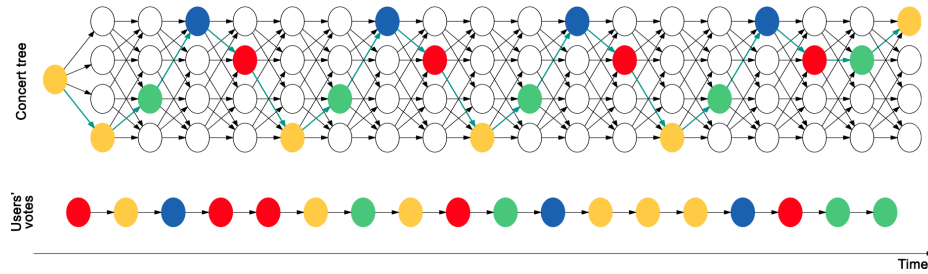
Fig. 9. Comparison between concert tree and users' votes.

5   Conclusion and Future Works

In this paper we presented a system for conveying audience emotions to orchestra musicians through a humanoid robot, and we described its architecture along with a brief overview of the algorithms we used. We have also discussed the first results collected after the deployment of our system in a real scenario, in particular by evaluating users' opinions during and after a live musical exhibition.

The results are encouraging for further studies. In particular, we noted that most of the users enjoyed the concert, and noted that the robot may have influenced their opinions. While this observation seems encouraging, further investigations are required to strengthen this hypothesis. To this end, we are planning to deploy our system in another live musical exhibition, and to compare it with other solutions, in order to understand the differences between humanoid robots and other kind of hardware in terms of their ability to convey emotions.

6   Acknowledgements

References

1. Anzalone, S.M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.L., Bodeau, N., Maharatna, K., Chetouani, M., Cohen, D., Group, M.S., et al.: How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot. Research in Autism Spectrum Disorders 8(7), 814–826 (2014)
2. Anzalone, S., Cinquegrani, F., Sorbello, R., Chella, A.: An emotional humanoid partner. In: Proceedings of the 1st International Symposium on Linguistic and Cognitive Approaches to Dialog Agents - A Symposium at the AISB 2010 Convention. pp. 1–6 (2010)
3. Augello, A., Infantino, I., Pilato, G., Rizzo, R., Vella, F.: Binding representational spaces of colors and emotions for creativity. Biologically Inspired Cognitive Architectures 5, 64 – 71 (2013)
4. Brown, L., Howard, A.M.: Gestural behavioral implementation on a humanoid robotic platform for effective social interaction. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. pp. 471–476 (2014)
5. Burger, B., Bresin, R.: Communication of musical expression by means of mobile robot gestures. Journal on Multimodal User Interfaces 3(1), 109–118 (2010)
6. Chella, A., Sorbello, R., Pilato, G., Vassallo, G., Balistreri, G., Giardina, M.: An architecture with a mobile phone interface for the interaction of a human with a humanoid robot expressing emotions and personality. In: Congress of the Italian Association for Artificial Intelligence. pp. 117–126. Springer (2011)

7. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Commun. 40(1-2), 5–32 (2003)
8. Ekman, P.: Basic Emotions, pp. 45–60. John Wiley & Sons, Ltd (2005)
9. Gentile, V., Sorce, S., Gentile, A.: Continuous hand openness detection using a kinect-like device. In: 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems. pp. 553–557 (2014)
10. Gentile, V., Sorce, S., Malizia, A., Gentile, A.: Gesture recognition using low-cost devices: Techniques, applications, perspectives [Riconoscimento di gesti mediante dispositivi a basso costo: Tecniche, applicazioni, prospettive]. Mondo Digitale 15(63), 161–169 (2016)
11. Gentile, V., Milazzo, F., Sorce, S., Gentile, A., Pilato, G., Augello, A.: Body Gestures and Spoken Sentences: a Novel Approach for Revealing User's Emotions. In: Proceedings of 11th International Conference on Semantic Computing (IEEE ICSC 2017) (2017)
12. Hoffman, G., Bauman, S., Vanunu, K.: Robotic experience companionship in music listening and video watching. Personal and Ubiquitous Computing 20(1), 51–63 (2016)
13. Lim, A., Ogata, T., Okuno, H.G.: Towards expressive musical robots: a cross-modal framework for emotional gesture, voice and music. EURASIP Journal on Audio, Speech, and Music Processing 2012(1), 3 (2012)
14. McCallum, L., McOwan, P.W.: Face the music and glance: How nonverbal behaviour aids human robot relationships based in music. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. pp. 237–244. HRI '15, ACM, New York, NY, USA (2015)
15. Meudt, S., Schmidt-Wack, M., Honold, F., Schüssel, F., Weber, M., Schwenker, F., Palm, G.: Going Further in Affective Computing: How Emotion Recognition Can Improve Adaptive User Interaction, pp. 73–103. Springer International Publishing, Cham (2016)
16. Posner, J., Russel, J.A., Peterson, B.S.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology 17(3), 715–734 (2005)
17. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
18. Soleymani, M., Caro, M.N., Schmidt, E.M., Sha, C.Y., Yang, Y.H.: 1000 songs for emotional analysis of music. In: Proceedings of the 2Nd ACM International Workshop on Crowdsourcing for Multimedia. pp. 1–6. CrowdMM '13, ACM, New York, NY, USA (2013)
19. Sorbello, R., Chella, A., Calí, C., Giardina, M., Nishio, S., Ishiguro, H.: Telenoid android robot as an embodied perceptual social regulation medium engaging natural human-humanoid interaction. Robotics and Autonomous Systems 62(9), 1329 – 1341 (2014), intelligent Autonomous Systems
20. Sorbello, R., Chella, A., Giardina, M., Nishio, S., Ishiguro, H.: An architecture for telenoid robot as empathic conversational android companion for elderly people. In: Intelligent Autonomous Systems 13, pp. 939–953. Springer (2016)
21. Spataro, R., Chella, A., Allison, B., Giardina, M., Sorbello, R., Tramonte, S., Guger, C., La Bella, V.: Reaching and grasping a glass of water by locked-in als patients through a bci-controlled humanoid robot. Frontiers in Human Neuroscience 11, 68 (2017)
22. Tkalčič, M., De Carolis, B., de Gemmis, M., Odić, A., Košir, A.: Introduction to Emotions and Personality in Personalized Systems, pp. 3–11. Springer International Publishing, Cham (2016)