



# Analysis of Affective Behavior in the Artistic Installation Moviescape

Elena Partesotti<sup>1</sup>✉, Eduardo D. Hebling<sup>1</sup>, Artemis S. Moroni<sup>2</sup>, Micael Antunes<sup>1</sup>, César P. da Silva<sup>1</sup>, Cássio G. Dezotti<sup>2</sup>, and Jônatas Manzolli<sup>1</sup>

<sup>1</sup> Institute of Arts/Interdisciplinary Nucleus for Sound Studies, University of Campinas, Campinas, Brazil

[eparteso@unicamp.br](mailto:eparteso@unicamp.br)

<sup>2</sup> ICTI Renato Archer and NRVC, University of Campinas, Campinas, Brazil

**Abstract.** The purpose of this paper is to study the correlation between attention and affective response when a user interacts with the artistic installation MovieScape - an Extended Immersive Digital Music Instrument. We analyzed the affective modulation in ten subjects. For that, we applied Affective Slider and audio-segmentation on the recorded performances of a Circumplex Space (a 2D Circumplex Model of Affect) to apply quantitative metrics and divide the interaction in Explorative and Contemplative states. In our paper we propose this as a unifying conceptual framework for analysing affective behavior within interactive installation such as MS. Overall, the analyses present an increase of Arousal after the interaction of the subjects within the environment, and a correlation between the Explorative interaction and the increment of Arousal which confirms the improvement in attention. Eventually, an inductive mechanism of expectancy occurred. These conclusions open up for potential in designing EIDMIs for therapeutic and pedagogical purposes. Also, these findings may also prove that subjects reached Creative Empowerment while becoming aware of MovieScape.

**Keywords:** EIDMI · Affective Modulation · Learning · Creative Empowerment · Early Cinema

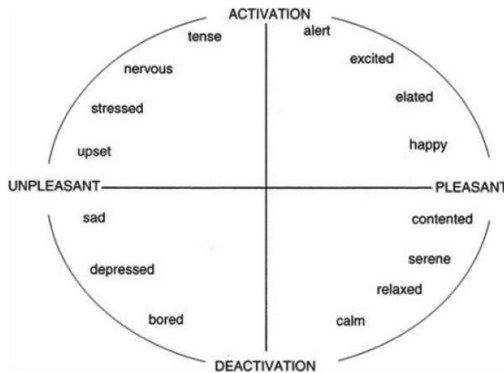
## 1 Analysis of Affective Behaviour in the Artistic Installation Moviescape

### 1.1 Introduction

Research on the paradigm of Embodied Cognition is still under discussion with various approaches and proposals (Fausto Caruana and Anna M. Borghi 2013). This approach considers that cognition is a highly embodied or situated activity, and suggests that active beings ought to be considered first and foremost as acting beings (Anderson 2003). Mind and body are not treated as separate entities anymore but rather recognised as closely influencing one another, intermingled with each other.

Our research is based on emotion recognition, through a new unifying conceptual framework, and the application of audio-visual modalities as driving components in examining the affective behavior of the user. This unifying conceptual framework is based on an immersive Digital Musical Instrument in which audiovisual components are intertwined with the bodily movements of the user. In this situation, it is possible to understand the affective behavior through the mechanism of reinforcement learning (i.e. user's experience influences subsequent choices). Hence, in this paper we couple real-time interaction with video and music to verify through an initial pilot study how this environment induces affective responses in the users. Also, we designed the users' interaction based on body movements within the environment and proposed a virtual affordance that induces an active body response.

The Fig. 1 shows a representation of the Circumplex model mentioned above. We introduced the notion of the "Circumplex Space" associated with the Circumplex Model, and it allowed us to represent and analyse the collected data in a two-dimensional space. Since modulations in affective states can be represented as displacements in a Circumplex Space, we applied quantitative metrics to support our analysis.



**Fig. 1.** The Circumplex Model of Affect. Note: A graphical representation of the Circumplex Model of Affect with the horizontal axis representing the Valence dimension and the vertical axis representing the Arousal or activation dimension (Posner et al. 2005).

## 1.2 A Multimodal Installation

The emotion recognition through body gesture is still a topic of debate, and is based on a pre-existing dataset that relies mainly on body movement and face recognition, which provides qualitative and quantitative observations of the investigated subjects. Noroozi et al. (2018) presented an exhaustive survey on this topic, with a ready-made database, in which extraction of features was made with machine learning methods.

The authors defined a general pipeline of methods and eventually highlighted the difficulty in detecting general patterns of affective body language because these are mainly based on naive geometrical representations. Thus they described three main categories that do not consider sounds. However we are proposing a new technique, in

which the user is placed within a Digital Musical Instrument (DMI) of the two modalities - visual and sound - to verify which interaction is generated between these two as an affective result, and most of all to analyse and observe the possible application of this technique in the therapeutic and pedagogic domain. Our framework is based on the affective recognition with visuals and in particular with sounds that are generated in real time through user's behaviour in the space. A DMI is a now widespread label that represents a large whole, within which different types of technological interfaces can be found. In this paper we are referring to a kind of DMIs in which the subject interacts straightly through her body gestures within a space, considering then that the instrument become an Extended and immersive Digital Music Instruments (Partesotti, 2023, forthcoming).

MovieScope (MS), the artistic installation we propose, hence includes visual and sound multimodal interaction, in which the musical modality is a sonification algorithm that generates sounds according to the subject's movement (Table 1) shaping the affective parameter by the learning experience that the subject has of the two modalities. Consequently, it is not an issue of predetermined data, but cues that produce a self-organised affective relationship in real-time. A multimodal installation such as MS can be understood either as immersive infrastructure or interactive installation. The first could interact, generate, analyse and store multimodal information such as audio and video, among others (Manzolini 2015). The latter can be defined as facilitating both the physical and emotional engagement of the audience, involving body interaction in the reconfigured space of the art-related context (Jeon et al. 2019; Nam and Nitsche 2013).

In general, multimodal installations can help to investigate whereby the creative process takes place through interactive techniques based on the integration of multimodal signals and the study of the man-machine interplay. In this context, the observation of the user's Creative Empowerment (Partesotti et al. 2018) - a concept described further on - is embedded through the variation of the affective modulation that occurs within the multimodal interactive installation MS. In this perspective, as already pointed out by Partiesotti et al. (2018), we emphasise again the need to consider the paradigm of Embodied Cognition when designing general DMIs for individuals. As highlighted by Tahiroğlu (2021) during the embedded practice with a Extended Immersive DMI (EIDMI), they became producers, builders and composers. Since MS is a multimodal artistic installation controlled by the user's body performance, we decided to run a preliminary study to test how combined modalities' perception can induce affective response in the users. An example that has been used for designing MS is ADA Intelligent Space representing a synthetic organism (Wassermann et al. 2003). This space was implemented to test the interaction between "the stimuli generated by the space and users' behaviours as a way of producing synthetic emotions and their expression in sound material" (Wassermann et al. 2003, p. 395). In other words, it proposed a modulation through the interaction with the public. With the same analytical model with which ADA was programmed-working with an association between interactive sonification and users' behaviour-presented in the 2019 Artech Conference (Hebling et al. 2019) MovieScope as multimodal installation inspired by the experiential environment and the poetics of silent early cinema.

### 1.2.1 Silent Cinema Historical Frame

Former theories established silent cinema as a primitive and preparatory step for sound cinema, since its representation modes were not fully defined but in constant evolution. Silent cinema is now recognised as a discontinuous time, with two main periods: early cinema with its exploratory participation and dynamic investigation (from 1895 to 1915), and the silent narrative cinema with its well-defined narrative codes (from 1915 to 1929).

Early cinema was based on the “primitive mode of representation” (Burch 1990), and in the “aesthetic of astonishment” (Gunning 1995) which means presenting a series of discontinuous situations to cause a shock on the viewers. The public had active and exploratory participation in the showing by contributing to the sound environment - i.e. by singing, commenting, talking- as “additive to the spectacle” (Châteauevert and Gaudreault 2001). These perspectives were paramount during MS design. The films were presented in amusement parks, fairs, festivals, salons - among others - (Abel 2010; Abel and Altman 2001; Altman 2004; Gaudreault et al. 2012), where the audience could have an exploratory approach towards the film experience. The films could be accompanied by storytellers, voice actors, sound effects artists or machines, and also musicians (Altman 2004). Later on, the films begin to be shown in small storefront theatres, called nickelodeons (Musser 1983).

In these places, the sound of the spectators and the chaos (Altman 2004) generated by the many accompaniments above mentioned, began to organise themselves. Cinema entered its “institutional mode of representation” (Burch 1990) - as we know it today -, since it is already an industrialised system of production and distribution of films, characterised by the supremacy of Hollywood (after 1915). This is the “cinema of narration” (Gaudreault 1987) where the public doesn’t participate actively in the sound environment as it has been silenced by the institutionalised system (Châteauevert and Gaudreault 2001). Music, from now on, became the only mode of accompaniment and more standardised due to musical publications called photoplay, musical suggestions or mood music (Altman 2004; Hebling 2017; Sauer 1998). After its standardisation, music grew more complex to strengthen the narrative structure and the psychological traits of the movie’s characters. As Gorbman (1987, p. 6) states, music had a function of a “bath of affect”, i.e. the ability that film music has to induce the viewer into a specific psychological state, using “the harmonic, rhythmic, melodic suggestiveness and channeling effects of music.”

### 1.3 MovieScape, a Multimodal Installation

Literally inspired by this changing – from active participation to passive assimilation of the audience – and the self-organising environment, in 2019 we designed the multimodal installation MS emulating the nickelodeon environment in an artistic context. We made thus the public relive the curiosity, uncertainty, and exploration offered by that new environment, through musical stimuli, environmental noises, and the moving images novelty.

MS has been designed to interactively generate sounds induced by the trajectory of the participant’s body in the room. Once into this environment, the participant sees the projection of a large sphere surfaced by a number of windows linked to scenes from a

silent film (in this case *Aurora*, Murnau 1927), as shown in Fig. 3b. Previously informed, the user grabbed a virtual steering wheel with all the yoke's movement possibilities. That is to say, the user grabs a virtual steering wheel to activate the MS environment, but eventually she has to use yoke's directions in order to control it. In doing so she approaches the sphere and selects a window that expands and the film's scene, with its relative musical theme, becomes a moment of cinematic viewing, thus triggering the projection of the chosen scene and its predefined music. Hence, she can move between an Exploratory interaction, with environmental sounds and film windows, to a Contemplative one, with musical themes and complete film scenes. This is an emulation of the experience that took place on the nickelodeon stages, and also a metaphor for the historical transition between early cinema and narrative silent cinema (more on this in p. 11). The interplay of movement and sound, guided by the interactive video, can be a fruitful experience to understand how new approaches for audiovisual interactions could be designed.

#### 1.4 Purpose of the Study and Methodology

Our aim is to contribute to the field with our multimodal installation MS and verifying that the nature of interaction is connected to the concept of Creative Empowerment (Partesotti et al. 2018). Hence, to verify that MS enables users to reach Creative Empowerment we analyse the affective modulation in ten subjects through a qualitative and quantitative approach. We retrieved data on Valence and Arousal of the subjects through Affective Sliders (AS) (Betella and Verschure 2016) and recorded their performance in order to do an audio-segmentation of each interaction. To be in control over something the user is doing, means that she is able to continue or determine that situation at will (McClelland 2020). Thus, Creative Empowerment (CE) accounts for the rewarding experience and the perception that occurs while gaining full control over our actions during the interaction itself (Partesotti et al. 2018). Given the aforementioned condition, we believe that with the multimodal installation MS is possible to observe and validate that CE is correlated with the self-perceived reward outcome. This would be possible through the variation of Valence and Arousal collected with AS and the audio segmentation of each user's performance. Moreover, we assume that CE is a process that increases the self-esteem, attention, and satisfaction of the user. The importance of this hypothesis and the results we obtained concerns the applicability in the therapeutic and pedagogical fields, in which a creative perspective may support learning a goal. Interactive environments that implement music have proved their benefits in the treatment of diverse pathologies and education. For example in the relief of stress (Taneja et al. 2017) and improvement in mental state and learning desire of students (Kanehira et al. 2018), or in learning of cognitive, motor, psychological, social and to stimulate musicality (Correa et al. 2009). This falls within the domain of Reinforcement Learning which studies behavior through repeated trial-and-error interactions within a dynamic environment. In this domain, user's experience alters the estimates for future rewards influencing subsequent choices (Lee et al. 2012). People continually make choices based on prediction (good for me - bad for me), that is to say that individuals learn by interacting with the environment (Sutton and Barto 2018).

MS has been performed at NICS Studio - where it was originally designed - and proposed to ten volunteers. We analysed their performances to track their behaviours

with the environment. To collect data we used AS (Betella and Verschure 2016), a validated visual scale tool where each participant self-reports her own current affective experience in a two-dimensional model. We collected them before and after volunteer interaction, observing their change in Arousal and Valence according to the Circumplex Model of Affect (Posner et al. 2005) as well as video recordings and questionnaires.

### 1.5 Architecture of MS

Indeed, this environment contains a virtual part (audiovisual), but is closely linked to user performance which is critical for determining the type of interaction in the multi-modal installation (Partridge and Rowe 1994). This approach is useful for the reader’s understanding of the type of interaction that results after proposing the environment to ten subjects, that is to say, the nature of the relationship between the user and the environment. An audiovisual scape is selected with an imaginary steering wheel when the user closes both fists. After this initial gesture, a sphere covered with images of the silent cinema is modified according to the movement of the participant; the wheel rotations alter the succession of scenes, while the displacement of the subject affects the sonification.

This steering wheel, as previously mentioned, can work not only to turn right or left, but it can also direct the moves up or down. For this reason, we refer to this steering wheel as a yoke of an airplane.

Figure 2a) shows a brief comparison with the movements of an aircraft. We used the movements of Yaw and Pitch. Figure 2b) Image showing the negative movements mapped and the result on the sphere. Moving the steering wheel upwards (red) makes

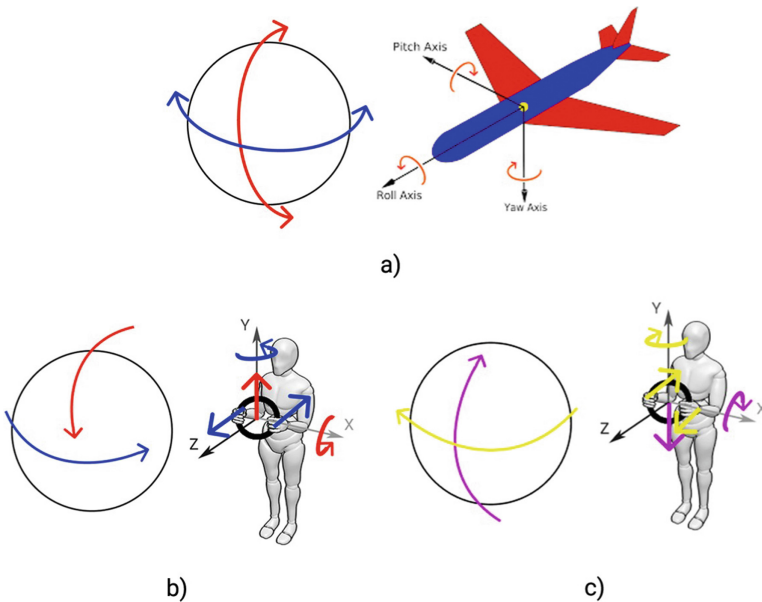


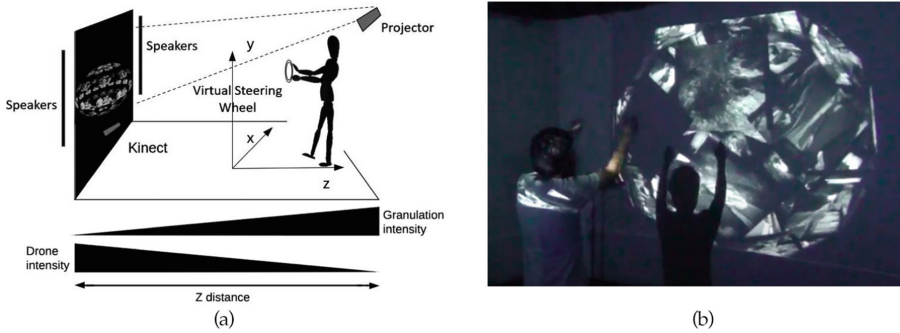
Fig. 2. a, b, c: Yoke

the sphere rotate around the X axis. Rotate steering wheel, with right hand forward and left backward (blue), in z axis, rotates the sphere around the Y axis. Figure 3c) Image showing the positive movements mapped and the result on the sphere. Moving the steering wheel downwards (purple) makes the sphere rotate around the X axis. Rotate steering wheel, with left hand forward and right backward (yellow), in z axis, rotates the sphere around the Y axis.

**Table 1.** MovieScape Gesture Table

Gesture name	Gesture definition	Result
Grab Virtual Steering Wheel.	Close both hands with arms extended in front of you.	Start manipulating the sphere.
Steering Wheel up.	Move both arms with hands closed upward in the Y axis.	Make the sphere rotate positive around the X axis.
Steering Wheel down.	Move both arms with hands closed downward in the Y axis.	Make the sphere rotate negative around the X axis.
Steering Wheel front.	Move both arms with hands closed frontward in the Z axis.	Make the sphere zoom in.
Steering Wheel back.	Move both arms with hands closed backward in the Z axis.	Make the sphere zoom out.
Turn Steering Wheel right.	With hands closed, right hand forward and left backward in the Z axis.	Make the sphere rotate positive around the Y axis.
Turn Steering Wheel left.	With hands closed, left hand forward and right backward in Z axis.	Make the sphere rotate negative around the Y axis.
Move forward in the room.	Move the body forward in the Z axis.	Increases the Drone volume, decreases the grain volume and decreases the grain size.
Move backward in the room.	Move the body backward in the Z axis.	Decreases the Drone volume, Increases the grain volume and Increases the grain size.
Move right in the room.	Move the body to the right in the X axis.	Increases the delay and the rarefaction.
Move left in the room.	Move the body to the left in the X axis.	Decreases the delay and the rarefaction.

The sonification is spatialized so that the position of the user relative to the room controls the sound. As the participant moves away from the projection screen, the granularity of the sound increases and the soundtrack becomes noisier. As the user approaches the projection screen, a drone sound (i.e. a low frequency sound texture) intensifies (Fig. 3b shows a participant interacting with MS). Since our analysis is mainly grounded on the



**Fig. 3.** The MovieScape installation

sonification factors, for the sake of the interpretations we divided the type of interaction of the users into two main categories: Contemplative and Explorative, which we will discuss later on.

The design process of MS was anchored in a participative methodology that is a factor of knowledge construction and artistic expression, a process elsewhere named cultural sharing (Çamcı and Hamilton 2020). Figure 3 shows the basic setup of MS with the projection screen, projector, motion sensor, sound monitors, and the subject manipulating a virtual yoke. In it, the coordinates  $X$ ,  $Y$ ,  $Z$  represents the movement in space, and the location  $(0,0,0)$  is in the centre of the installation. For more technical details see Hebling et al. (2019).

In MS we have designed a multimodal interface where the interaction occurs with an imaginary steering wheel and therefore offering precise Virtual Affordances to the subject. This Virtual Affordance can also referred to as Multiple Trajectories (Partesotti et al. 2018). This interaction, moreover, recalls a yoke controller that is activated when the user closes her two fists, “driving” the sphere and thus navigating it while observing the distributed movie segments. This allows the adjustment of the images and the rotation of the soundscapes and setting the audiovisual scenes to create a narrative progression over time. In fact, this approach of ours derives from the need to allow the user to become the protagonist who can combine scenes and therefore edit the film by herself.

## 1.6 Sonification

As mentioned above, in MS we divided the type of interaction of the subjects into two main categories: Contemplative and Explorative. This categorisation is grounded on a technical and interpretative understanding of the experience of the subject with diverse audiovisual materials.

We chose to underline the connection existing between the inductive experience and the affective one. From the technical perspective, the Explorative interaction generates two different sound layers: a granular synthesis with a continuous texture (called drone), and noisy sounds (see Fig. 3a). The drone sound is inspired by empty ambient sounds; the noisy sounds are inspired by the cinema environment, i.e. the sound of the projector, door, people talking and chair cheeks. The MS dynamic design is mapped in the position

of the subject in the room. The farther the subject is from the screen, the louder the noisy sounds. When the subject goes far from the screen, drone sound volume goes down, and noisy sounds volume goes up. On the contrary, when the subject gets close to the screen, drone sound volume goes up, and noisy sounds volume goes down; furthermore, the noisy sounds can be heard whenever a window is chosen (Hebling et al. 2019). The Contemplative interaction is characterised by matching a scene with its music theme. The musical themes environment is based on the set of themes chosen from, or inspired by musical suggestions books (Zamecnik 1913a, 1913b, 1914), and recorded with a piano. This moment is inspired by the public settling in a more immersive state, induced by musical themes in a Gorbman's "bath of affect" (1987) (see p. 6). As for the interpretative perspective of a user's experience with MS, during the Explorative moment there is a higher level of body activity associated with anticipation, as the inductive mechanism underlines (Huron 2006). That is to say, the user's physical activity rises and coincides with an increase in arousal.

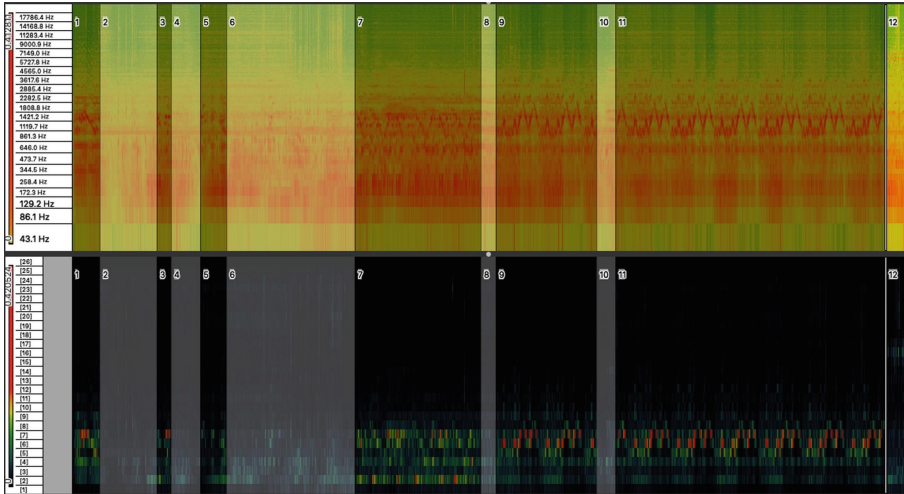
### 1.7 The Experiment

The pilot test was conducted using 10 participants from the University of Campinas (Unicamp). We asked the subjects to enter the NICS' studio - a room with MS running. Once in the room each participant had up to five minutes to interact with the installation, while the whole interaction was video-recorded.

A video camera, in fact, was positioned so that it was possible to capture the participant's movements and the entire wall on which the images were displayed. Considering that the subjects and the dataset concern the type of multimodal interaction that occurs during the interaction, we analyse the behaviour and codetermination of the subject in the environment concerning the audio-visual feedback provided by the environment itself. To do this, we used Affective Slider (Betella and Verschure 2016) before and after the subjects' performance to verify the changes in Valence and Arousal and analysed the relationship with movement and affective response using the sound and visual feedback parameter. In a pre-analysis stage, all the AS reports were measured and plotted on a scale from 1 to 1. To obtain the information about the interactions with MS, audio segmentation was applied, collecting data on the time each participant spent in the Explorative and Contemplative interaction and the number of variations. We based our analyses in two datasets: state data - the time each subject stays in the Explorative and Contemplative state and Circumplex data - the level of Arousal and Valence of each subject before and after the experiment. We describe our methodology of analysis in three steps: 1. Data structuring, 2. Quantitative Measurements, 3. Data crossing. Furthermore, considering that our sound database is generated by an algorithm, it denotes an interactive relationship in real-time.

Data Structuring: In that step, we structured the data to perform the analysis. To structure the state data, we performed the audio segmentation of each experiment session. The goal of the segmentation of the sound files is to split the granular synthesis and the musical theme of the environments experienced by the subject. To do that, we applied two audio descriptors of Sonic Visualizer (Cannam et al. 2010): the spectrogram and the bark coefficients (Bullock 2007). The segmentation is performed by observing the dispersion of the spectral energy in the spectrogram and in the bark coefficients, aligned

with the listening of the audio file. In Fig. 4 we show an example of audio segmentation with Sonic Visualizer. The odd numbers represent the granular synthesis (Explorative) and the even numbers represent the musical themes (Contemplative).



**Fig. 4.** Segmentation of the audio of the experiment. Above, the spectrogram, and below the bark coefficients descriptors.

**Quantitative Measurements:** In that step, we performed quantitative measurements in the two data sets. In the state data, we summarise, for each participant, the time of Explorative and Contemplative state. Then we reached the percentage of the two states for each participant. To perform the quantitative measurements of the Circumplex data, we used the Circumplex Space. By observing the points in the Circumplex Space before and after the experiment we calculated: a) the trajectory and general tendency of each participant in the Circumplex Space; b) the magnitude of the change of each participant in the Circumplex Space; c) the percentage of change in Arousal and Valence. **Data crossing:** Aiming to understand how MS modulates the affective states of the users, we crossed the State and Circumplex data. Our analytical guideline is that the total time in different states affords different levels of change in the Circumplex Space.

## 2 Results

As for the analysis, we represented the Circumplex Model of Affect in a 2D Circumplex Space, with Arousal and Valence representing the affective states. The results are plotted in the following graphics. Figure 5 presents how the affective experiences had evolved for each participant. Of a total of ten participants, five increased both their Arousal and Valence, four increased Arousal and decreased Valence, while the last displayed a decrease in both of them.

Figure 5 shows how the affective states of each participant evolved in the Circumplex Space; the y-axis represents Arousal and the x-axis represents Valence.

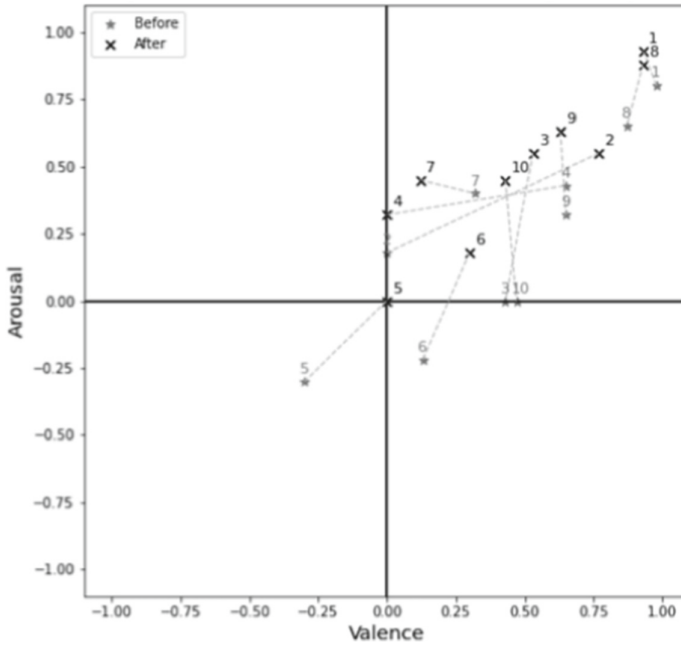


Fig. 5. Projection of the Results in the Circumplex Space

The levels of Arousal before and after the experiment can be observed in Fig. 6. In the figure, the y-axis represents Arousal. The x-axis presents the total time each participant spent in the Explorative state during five minutes of interaction. The remaining time was spent in the Contemplative state.

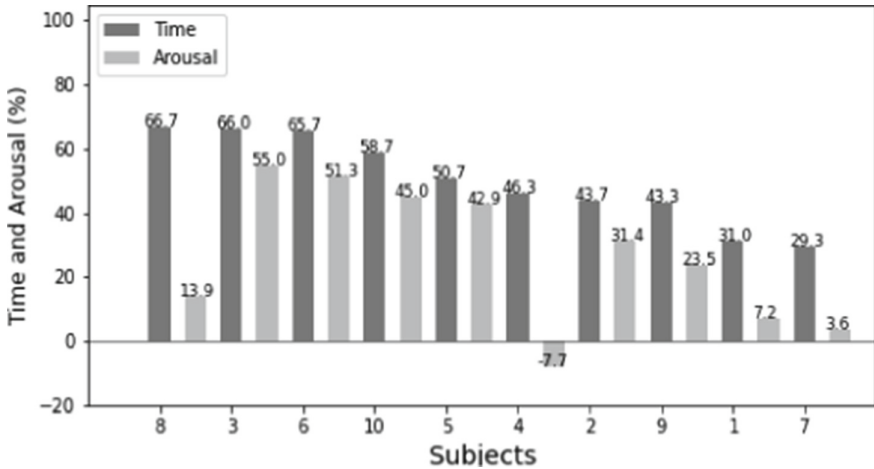


Fig. 6. Arousal Before and After the Experiment.

Figure 7 shows the percentage of Arousal variation and the percentage of the time (within 5 min) which was spent in the Explorative interaction. The remaining time was spent in the Contemplative interaction. It was observed that, in general, the more time the participants spent in the Explorative interaction, the more the Arousal increased. To highlight that observation, we organized the subjects in the x-axis in a decrescent order of time spent in the Explorative interaction, from left to right. The Arousal columns also are decrescent, except by two outliers, subjects 8 and 4. Subject 4 related in the form some discomfort to interact with MS and that probably is related to the low Arousal response. Subject 8 didn't give any cues in the form which would contribute to the analyses.

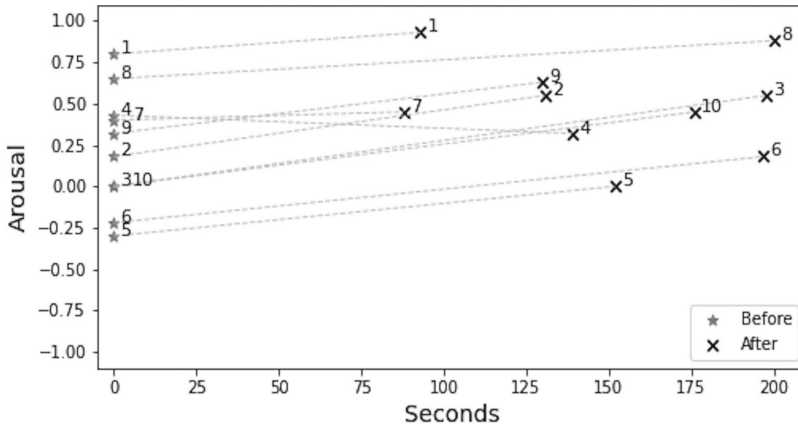


Fig. 7. Percentage of Arousal Variation and time in Explorative state.

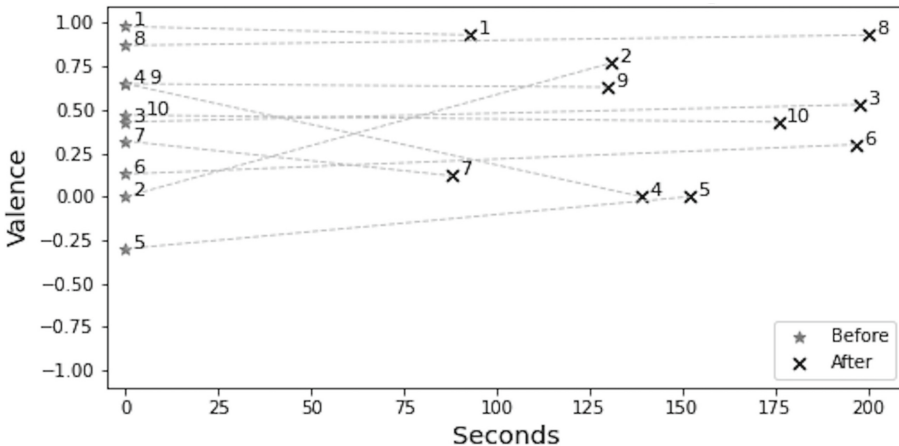


Fig. 8. Valence Before and After the Experiment.

By examining the Explorative and Contemplative time spent by the subjects, we conclude that there is a correlation between the Explorative interaction and the increment of Arousal.

In Fig. 8 the y-axis represents Valence. The x-axis presents the total time each participant spent in the exploratory state during five minutes of interaction. The remaining time was spent in the Contemplative state.

Figure 9 shows the percentage of Valence variation. X-axis shows the ranking of the time percentage that each participant spent in the Explorative state. Y-axis shows the variation of the time within 5 min which was spent in the Explorative state. The remaining time was spent in the Contemplative state.

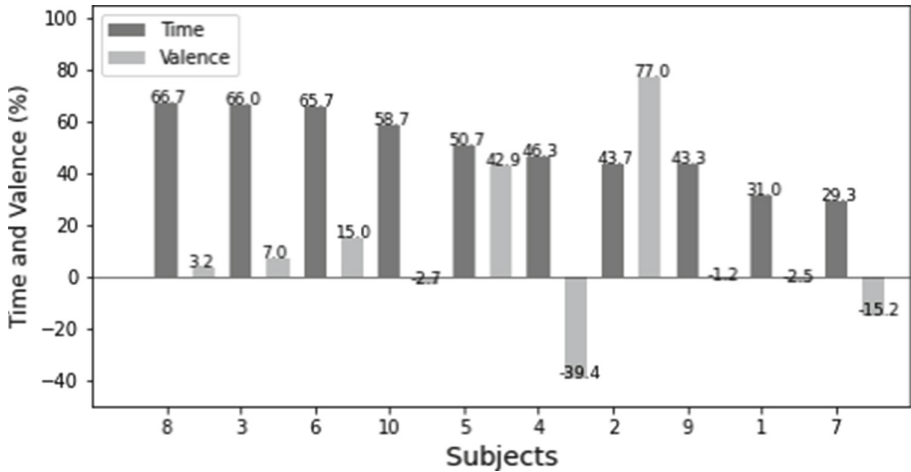
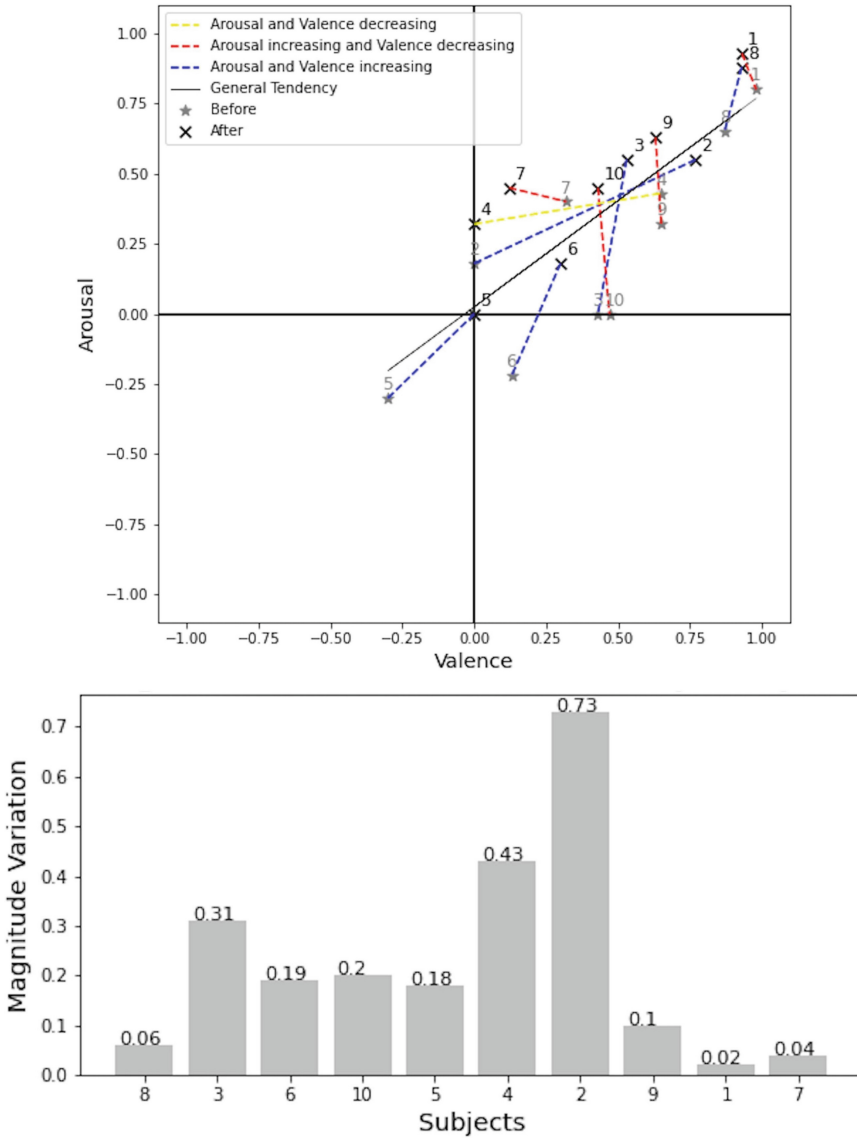


Fig. 9. Percentage of Valence Variation.

Figure 10 displays how the affective states of each participant evolved in the Circumplex Space. The black line represents the trend of the affective states reported in the experiment and obtained through linear regression. Blue vectors indicate Arousal and Valence increase, red vectors indicate Arousal increase and Valence decrease and yellow vectors indicate Arousal and Valence decreasing. We can infer that the MS experience led to a general tendency of increment in Arousal levels.



**Fig. 10.** (up) and (Below). Up tendency of results in the Circumplex Space. (Below) Representation of the magnitudes of the changes in the Circumplex Space, considering the points before and after the experiment for each participant.

### 3 Discussion

We presented the quantitative analyses of two parameters. First, the levels of Valence and Arousal before and after the user’s interaction following the AS questionnaire, and second, the type of interaction according to the audio segmentation of each user, considered

as Explorative and Contemplative. From the graphics of Valence and Arousal, it emerges that users have learned how to use ‘the metaphor of driving’ that we designed, showing an increase in attention. We have also observed that MS strengthens the interaction between the environment and the cognizer, allowing an enactive experience. Furthermore, as pointed out by Huron (2006), Arousal increases when the object of interaction generates expectation. In this context, the analysis of the users and their performance showed an increment in the level of expectation which corresponds with an increase in Arousal. This strengthens our position that MS can produce a sense of CE associated with an increased level of Arousal.

From the analysis, we observed that once the subjects became aware of the virtual affordance (section II-C) in order to interact with MS, then their behavior improved and their willingness to interact with the environment increased more, too. This is represented by a general increase in Arousal level. Interacting with MS’s environment, the user could reach the state of CE in which she enjoys the experience while showing an increased level of attention and satisfaction. Indeed, MS could allow us to observe that CE is associated with the self-perceived reward outcome. CE leads to a co-determination process (Partesotti et al. 2018) and displays important changes which may be useful in future studies at an interdisciplinary level. These findings confirm how theoretical approaches grounded on the paradigm of Embodied Cognition are strategic to the design of an interactive installation which may find application in cognate fields.

As a general observation, the participants of the studied population were led towards an Arousal increase. Hence, as an analytical result, we observed that an inductive mechanism of expectancy occurred within MS. In addition to it, from the interpretative perspective, we confirm that the increase in Arousal could correspond with the reaching of CE of the subjects. Considering the exposure effect described by Huron (2006, p. 135), familiar stimuli reduce the likelihood of orienting responses, which would suggest that familiarity would decrease Arousal and increase Valence. From the data collected from 10 subjects, 50% of the participants had an increase in Valence, and 50% had a decrease. If we consider the average, Valence has increased slightly. Indeed, Arousal increased significantly. Huron (2006) explains how expectation and familiarity are connected in a reward circuitry that modulates affect. According to his theory, expectancy leads to a raise of Arousal, since humans evolved in order not to get surprised. In opposition, familiarity leads to a decrease of Arousal, since “familiar stimuli reduce the likelihood of orienting responses, and that this reduces an organism’s Arousal level” (Huron 2006, p. 135). The author’s work is grounded in the context of tonal music, while in MS we extend his model to a multimodal environment. In fact, MS is a multimodal installation in which expectation’s outcome is linked to audiovisual and embodied stimuli, which may constitute a rewarding and immersive experience that in turn unfolds the empowerment of a creative expression. In this perspective we believe the visuals and sounds elements together can strengthen subjects’ reward process, and future studies should investigate this connection. As explained above, we categorise the subject’s interaction into two types: Explorative and Contemplative. As results show, the Explorative interaction is related to an increase of Arousal (Fig. 7), which we interpret as the expectancy of the interaction based on Huron’s theory. One central finding indicates that the subjects who

spent more time in the Explorative interaction were the ones presenting an augment of Arousal.

## 4 Conclusions

Most of the studies concerning emotion recognition, including the survey by Noroozi et al. (2018) apply a preset dataset primarily based on action-recognition. On the contrary, in our pilot study, we used a dataset as a result of a multimodal relationship, with special attention given to the pair sound-movement and in particular how sound feedback affects the emotions of the user. This pilot study made it possible to observe the analysis of emotion recognition from another perspective that is usually not considered, sowing the seeds for a possible methodology that can help in future analyses. Furthermore, this experiment allowed us to verify musical anticipation, and its potential, already widely described by Huron (2006), as an effective parameter for analyzing the emotional process of subjects within a multimodal interactive environment such as MS.

EIDMI are interfaces that can per se arouse creativity. With this insight, in this paper we have presented the artistic installation MS, designed under the paradigm of Embodied Cognition. Certainly, the opportunity provided by a multimodal environment may find applications also in other domains. For example, MS may be viewed as a multimodal interactive installation for a new modality of self-organization of movie material; this could be applicable in the pedagogic and therapeutic fields. In fact, as literature underlies (Hakvoort et al. 2020; O'Kelly et al. 2013; Pelletier 2004), in Music Therapy sessions the variation of Arousal distress, regulates attention and modulates emotional responses, empowering clients. Indeed, Music Therapy, a multimodal environment stimulating the sensory modalities and strengthening the CE process can be effective in the treatment of various health conditions (i.e. Autism Spectrum Disorder) (Berger 2002; Partesotti et al. 2018). In the educational field, MS could see its application within the historic and humanistic frame, while on therapy, it would be the case of action executions. The expectation of interactive environments such as MS applications is broad and can embrace a path directed on community improvement and social inclusion.

## 5 Future Directions

To improve the MS environment and to offer another sense of control, we intend to redesign the imaginary steering wheel (a bimanual yoke gesture control) to a unimanual gesture interface. It would be worthwhile investigating how CE will be affected by this alternative interface. In fact, since MS represents a first prototype of a multimodal artistic installation, our test has been a pilot study, too.

Further investigation over the role of expectation within a new version should be addressed with more volunteers. Moreover, further works and design of prototypes should bear in mind the 4E cognition paradigm (Ryan and Schiavio 2019) in order to contribute to a better understanding of the interaction that may occur between the cognizer and the multimodal environment. This call to interdisciplinarity can only generate new knowledge by opening a wider field in research. To conclude, Creative Empowerment is a key parameter as it is grounded on the expectation and reward circuits strongly linked

to an enactive experience. In other words, it could represent an innovative experimental framework based on the intervention of EIDMI in interdisciplinary contexts within the frame of reinforcement learning.

**Abbreviations.** The following abbreviations are used in this manuscript: AS: Affective Sliders, CE: Creative Empowerment, DMI: Digital Musical Instruments, EIDMI: Extended Immersive Digital Music Instruments, MS: MovieScape.

**Funding.** Elena Partesotti is supported by FAPESP, under Grant n. 2016/226190; Eduardo D. Hebling is supported by CAPES, under Grant n. 88887.313062/2019-00; Jônatas Manzolli is supported by CNPq projects, under Grant n. 429620/2018-7, 304431/2018-4; Micael Antunes is supported by FAPESP, under Grant n. 2019/09734-3.

## References

- Abel, R. (ed.): *Encyclopedia of Early Cinema*. Routledge, Abingdon (2010)
- Abel, R., Altman, R.: *The Sounds of Early Cinema* (1st ed.). Indiana University Press (2001). <https://doi.org/10.1353/not.2002.0123>
- Altman, R.: *Silent Film Sound*. Columbia University Press, New York (2004)
- Anderson, M.L.: Embodied cognition: a field guide. *Artif. Intell.* **149**(1), 91–130 (2003). [https://doi.org/10.1016/S0004-3702\(03\)00054-7](https://doi.org/10.1016/S0004-3702(03)00054-7)
- Berger, D.S.: *Music therapy, sensory integration and the autistic child* (5. impr, Vol. 22). Jessica Kingsley (2002)
- Betella, A., Verschure, P.F.M.J.: The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS ONE* **11**(2), 11 (2016). <https://doi.org/10.1371/journal.pone.0148037>
- Bullock, J.: Libxtract: a lightweight library for audio feature extraction. In: *Proceedings of the 2007 International Computer Music Conference*, no. 4 (2007)
- Burch, N.: *Life to Those Shadows*. University of California Press, Oakland (1990)
- Çamcı, A., Hamilton, R.: Audio-first VR: new perspectives on musical experiences in virtual environments. *J. New Music Res.* **49**(1), 1–7 (2020). <https://doi.org/10.1080/09298215.2019.1707234>
- Cannam, C., Landone, C., Sandler, M.: Sonic visualiser: an open source application for viewing, analysing, and annotating music audio files. In: *Proceedings of the 18th ACM International Conference on Multimedia*, p. 1467 (2010). <https://doi.org/10.1145/1873951.1874248>
- Châteauevert, J., Gaudreault, A.: The noises of spectators, or the spectator as additive to the spectacle. In: Abel, R., Altman, R. (Eds.), *The Sounds of Early Cinema* (1st ed., pp. 183–191). Indiana University Press (2001). <https://doi.org/10.1353/not.2002.0123>
- Correa, A.G.D., Ficheman, I.K., do Nascimento, M., de Deus Lopes, R.: Computer assisted music therapy: a case study of an augmented reality musical system for children with cerebral palsy rehabilitation. In: *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, pp. 218–220 (2009). <https://doi.org/10.1109/ICALT.2009.111>
- Etkin, A., Büchel, C., Gross, J.J.: The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**(11), 693–700 (2015). <https://doi.org/10.1038/nrn4044>
- Caruana, F., Borghi, A.M.: Embodied Cognition: Una nuova psicologia. *G. Italiano Di Psicol.* **1**, 23–48 (2013). <https://doi.org/10.1421/73973>
- Gaudreault, A.: Narration and monstration in the cinema. *J. Film Video Univ. Illinois Press* **39**(2), 29–36. JSTOR (1987). <https://www.jstor.org/stable/20687768>

- Gaudreault, A., Dulac, N., Hidalgo, S. (eds.): *A Companion to Early Cinema* (1st ed.). Wiley-Blackwell, Hoboken (2012)
- Gorbman, C.: *Unheard Melodies: Narrative Film Music* (1st ed.). BFI Pub - Indiana University Press, Bloomington (1987)
- Gunning, T.: The cinema of attractions: early cinema, its spectator, and the avant-garde. *Wide Angle* **8**(3–4), 63–70 (1986)
- Gunning, T.: An aesthetic of astonishment: early film and the incredulous spectator. In: Williams, L. (Ed.), *Viewing positions: Ways of seeing film* (Reprint from 1989 *Art&Text*, nos. 34, 1st, pp. 114–133). Rutgers University Press (1995)
- Hakvoort, L., de Jong, S., van de Ree, M., Kok, T., Macfarlane, C., de Haan, H.: Music therapy to regulate arousal and attention in patients with substance use disorder and posttraumatic stress disorder: a feasibility study. *J. Music Ther.* **57**(3), 353–378 (2020). <https://doi.org/10.1093/jmt/thaa007>
- Hebling, E.D.: As coletâneas musicais temáticas e os manuais: Uma introdução e um estudo de caso sobre os temas do exotismo no cinema mudo [Master, Universidade de Campinas UNICAMP] (2017). <http://repositorio.unicamp.br/jspui/handle/REPOSIP/330905>
- Hebling, E.D., et al.: MovieScape: audiovisual landscapes for silent movie: enactive experience in a multimodal installation. In: *Proceedings of the 9th International Conference on Digital and Interactive Arts*, pp. 1–7 (2019). <https://doi.org/10.1145/3359852.3359883>
- Huron, D.: *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, Cambridge (2006). [https://doi.org/10.1016/s0262-4079\(17\)31427-6](https://doi.org/10.1016/s0262-4079(17)31427-6)
- Jeon, M., Fiebrink, R., Edmonds, E.A., Herath, D.: From rituals to magic: interactive art and HCI of the past, present, and future. *Int. J. Hum. Comput. Stud.* **131**, 108–119 (2019). <https://doi.org/10.1016/j.ijhcs.2019.06.005>
- Kanehira, R., Ito, Y., Suzuki, M., Hideo, F.: Enhanced relaxation effect of music therapy with VR. In: *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 1374–1378 (2018). <https://doi.org/10.1109/FSKD.2018.8686951>
- Lee, D., Seo, H., Jung, M.W.: Neural basis of reinforcement learning and decision making. *Annu. Rev. Neurosci.* **35**(1), 287–308 (2012). <https://doi.org/10.1146/annurev-neuro-062111-150512>
- Manzollì, J.: Multimodal generative installations and the creation of new Art form based on interactivity narratives, pp. 32–45. In: *XVIII Generative Art Conference GA2015* (2015). <https://www.generativeart.com/>
- McClelland, T.: The mental affordance hypothesis. *Mind* **129**(514), 401–427 (2020). <https://doi.org/10.1093/mind/fzz036>
- Musser, C.: The nickelodeon era begins: establishing the framework for Hollywood’s mode of representation. *Framework J. Cinema Med.* **22**/23, 4–11 (1983). <https://doi.org/10.5040/9781838710170>
- Nam, H.Y., Nitsche, M.: Interactive installations as performance: Inspiration for HCI. In: *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction - TEI ‘14*, pp. 189–196 (2013). <https://doi.org/10.1145/2540930.2540976>
- Noroozi, F., Corneanu, C.A., Kamińska, D., Sapiński, T., Escalera, S., Anbarjafari, G.: Survey on Emotional Body Gesture Recognition (2018). ArXiv:1801.07481 [Cs]. <http://arxiv.org/abs/1801.07481>
- O’Kelly, J., James, L., Palaniappan, R., Taborin, J., Fachner, J., Magee, W.L.: Neurophysiological and behavioral responses to music therapy in vegetative and minimally conscious states. *Front. Hum. Neurosci.* **7** (2013). <https://doi.org/10.3389/fnhum.2013.00884>
- Partesotti, E., Peñalba, A., Manzollì, J.: Digital instruments and their uses in music therapy. *Nord. J. Music. Ther.* **27**(5), 399–418 (2018). <https://doi.org/10.1080/08098131.2018.1490919>

- Partesotti, E.: (forthcoming) Immersive Digital Music Instruments to Empower Wellbeing through Creativity in Creating Digitally. *Shifting Boundaries: Arts and Technologies - contemporary applications and concepts* (2023). Springer
- Partridge, D., Rowe, J.: *Computers and creativity*. Intellect (1994)
- Pelletier, C.L.: The effect of music on decreasing arousal due to stress: a meta-analysis. *J. Music Ther.* **41**(3), 192–214 (2004). <https://doi.org/10.1093/jmt/41.3.192>
- Peñalba, A., Valles, M.-J., Partiesotti, E., Sevillano, M.-Á., Castañón, R.: Accessibility and participation in the use of an inclusive musical instrument: the case of MotionComposer. *J. Music Technol. Educ.* **12**(1), 79–94 (2019). [https://doi.org/10.1386/jmte.12.1.79\\_1](https://doi.org/10.1386/jmte.12.1.79_1)
- Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **17**(03) (2005). <https://doi.org/10.1017/S0954579405050340>
- Ryan, K., Schiavio, A.: Extended musicking, extended mind, extended agency. Notes on the third wave. *New Ideas Psychol.* **55**, 8–17 (2019). <https://doi.org/10.1016/j.newideapsych.2019.03.001>
- Sauer, R.: *Photoplay music: a reusable repertory for silent film scoring, 1914–1929*. American Music Research Center - Colorado Boulder, pp. 55–76 (1998)
- Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction* (Second edition). The MIT Press, Cambridge (2018)
- Tahiroğlu, K.: Ever-shifting roles in building, composing and performing with digital musical instruments. *J. New Music Res.* **50**(2), 155–164 (2021). <https://doi.org/10.1080/09298215.2021.1900275>
- Taneja, A., Vishal, S.B., Mahesh, V., Geethanjali, B.: Virtual reality based neuro-rehabilitation for mental stress reduction. In: *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 1–5 (2017). <https://doi.org/10.1109/ICSCN.2017.8085665>
- Ward, D., Stapleton, M.: Es are good: cognition as enacted, embodied, embedded, affective and extended. In: Paglieri, F. (Ed.), *Advances in Consciousness Research*, vol. 86, pp. 89–104. John Benjamins Publishing Company (2012). <https://doi.org/10.1075/aicr.86.06war>
- Wassermann, K.C., Eng, K., Verschure, P.F.M.J., Manzolli, J.: Live soundscape composition based on synthetic emotions. *IEEE Multimed.* **10**(4), 82–90 (2003). <https://doi.org/10.1109/MMUL.2003.1237553>
- Zamecnik, J.S.: *Sam Fox Moving Picture Music vol. 1*. Sam Fox Publisher, Cleveland (1913a)
- Zamecnik, J.S.: *Sam Fox Moving Picture Music vol. 2*. Sam Fox Publisher, Cleveland (1913b)
- Zamecnik, J.S.: *Sam Fox Moving Picture Music vol. 3*. Sam Fox Publisher, Cleveland (1914)