

# Immersive environments, video tracking and collective interactivity on smartphone : a generic “dispositif”

---

Dominique Cunin<sup>1</sup>, Emmanuel Durand<sup>2</sup>,  
Michał Seta<sup>2</sup>

<sup>1</sup>École des Arts Décoratifs – PSL (EnsadLab), <sup>2</sup>Société des arts technologiques  
dominique.cunin@ensad.fr, edurand@sat.qc.ca, mseta@sat.qc.ca

## Abstract

In this article, we describe our joined research around building a technological pipeline suitable for creating artistic immersive installations that utilize a combination of computer vision and mobile phones as means for interaction, both in the sense of Human Computer Interaction (HCI) and interaction between human participants. We frame our research in certain historical and philosophical context and propose a generic but complex “dispositif,” or device, that has a potential for exploration of diverse artistic themes suitable for interactivity within immersive environments. Taking our technological stack as inspiration, we propose a theoretical and conceptual approach to such creation and offer a draft of possible taxonomy that we find useful in such contexts, namely *collaboration*, *cooperation* and *competition*. We also bring the theme of surveillance to limelight as a potential and valid theme for exploration that is based on the use of such technologies.

## Keywords

Real time sensorial immersion, Collective interactivity, interactive arts, mobile screens, surveillance technologies, camera-based pose detection.

## DOI

10.69564/ISEA2023-51-full-Cunin-et-al-Immersive-environments

## Introduction

---

Our research is situated in the context of contemporary artistic creation working with digital technologies. We cross several custom tools to compose a generic technical device offering a strong potential for aesthetic exploration. To that end, we interface three groups of technical solutions that address the co-situated, collective interactivity with the help of mobile phones, the sensory immersion via large-scaled video projections with sound and spectators detection via computer vision systems. The current degree of maturity of these tools makes it possible to experiment with effective artistic devices with the public, although some technological deficits still need to be resolved.

But this exploration requires artistic choices: what kind of statements can we uphold with these techniques? This question is at the root of this article, where we propose theoretical and conceptual foundations for a first taxonomy of collective interactivity made possible by our technical device. One of the goals of this article is to situate our work from a critical, theoretical and artistic point of view, which is why we expose our understanding of the major domains and notions that underlie our research. The notion of “dispositif” in the practice of interactive arts, considered as being two folded, allows us to make the link between technical considerations and artistic statements. Immersion is approached through its sensory stakes, essentially audiovisual, often spectacular in its capacity to trick our senses of perception and whose monumental scale is articulated in our pipeline to the over-individual scale of mobile screens.

Collective interactivity, which we define from theories belonging to the field of interactive arts, is compared to studies on collective interaction conducted in computer sciences and in HCI in particular. This allows us to clarify our own theoretical positioning from which we'll start our taxonomy attempt. This constitutes our main contribution to the research: to construct the conceptual foundations of a taxonomy whose ambition is to propose design guidelines in the creation of immersive devices of collective interactivity.

Thus, we do not propose here a taxonomy in the sense of a complete classification of existing works in order to organise them between each other according to their characteristics. On the contrary, we tend to anchor its origins in identified artistic intentions. This is why we begin by formulating a first list of key notions that define the modalities of inter-individual relations that can make situations of collective interactivity happen. Drawing

from these key notions, we suggest some imaginary examples that take into account the particularities of our generic apparatus (“dispositif”) in order to demonstrate its efficiency as a conceptual shifter leading to an artistic creation. Finally, because the spectators are both continuously observed by a tracking system, but at the same time are also able to act on the immersive projection using their smartphones, we draw inspiration from population surveillance systems as a guide for our artistic intentions in our future creations. A certain state of the art in the artistic, technological and societal fields allows us to root these intentions in a contemporary and political reality of the technologies that we manipulate (and that we also invent) in order to define a singular critical position that we want to support.

### Artistic “dispositif,” technical “dispositif”

---

In French, the term “dispositif” is commonly used to designate certain works of technological art. It can be translated in English by “device,” or “apparatus.” We'll use the French word “dispositif” in this text as the theories we are using in our research come from French-based works. It is probably in connection with its technical origin (cf Herme's n°25, *Le dispositif entre usage et concept*, 1999—review of the Institute of the sciences of the communication of the CNRS) that the “dispositif” came to designate a larger part of interactive art works, in expressions such as “artistic dispositif of collective interactivity.” The theoretical context of this use of the term in the field of technological art creation is well established and, while the details of its history are out of scope for this article, we need to recall here some part of it which will serve our purpose.

Beginning with Michel Foucault, Giorgio Agamben reminds us that a device is a network that can be established between a set of elements with varied origins and qualities: “speeches, institutions, architectural arrangements, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral, philanthropic proposals, in short: the said, as well as the unsaid.”<sup>13</sup> Furthermore, Agamben sets a vision of the world separated into two main categories, “the living beings (or substances), and the 'dispositifs' inside which they keep being seized”<sup>1</sup> (translation by the authors) and the relation between the beings and the devices would produce the birth of the subject. While an artistic device is heterogeneous, its components can be, nevertheless, identifiable: “More than a simple technical organisation, the “dispositifs”

puts into play different enunciating or figurative instances, engages institutional situations as well as processes of perception”<sup>11</sup> (translation by the authors).

The experience of being in the presence of, and being connecting with the work is the very prerequisite that enables the deployment of the “dispositif” to become an intelligible artwork that will produce sense. This experience is therefore also what motivates its creation. We could say that the “dispositifs” is what conditions the creation, and what regulates the spectator’s relationship to the artwork. We find here a very strong echo with the notion of interactivity that we could summarise as being a relationship appearing between a work and its spectators. This relationship is built from a network of interactions captured by physical peripherals and transmitted to a computer machine in order to be replayed by a software architecture (a set of programs) whose ambition is to structure the viewer’s experience in accordance with the work’s aesthetic intention. The way we understand the notion of interactivity seems to indicate that two poles exist inside the interactive “dispositif”: artistic and technical. These two poles influence each other reciprocally in what forms a coherent set of heterogeneous elements, a network of components of various natures that interlock the ones in the others to produce a single “dispositif,” similar to the one defined by Michel Foucault.

The artistic, or abstract, side of this “dispositif” implies the implementation of a conceptual strategy through the author’s aesthetic intention. For the artist, it is therefore a question of conceiving an aesthetic relation and consequently, of defining what one could call an aesthetic contract of relations. The technical side of the “dispositif” consists of both hardware and software. The hardware takes the form of a computer, or a network of computers, that are at the core of the architecture, to which other devices or computing machines can connect. The software architecture is made of computer programs that exploit the machine according to the artistic contract, and, finally, scenography in which the viewer/participant is considered to be part of the whole “dispositif.”

Through this idea of a two-fold “dispositif”, technical and artistic, we defend the idea that one technical device can support several artistic devices and, consequently, can be the starting point of the creation of several artistic works which will exploit the technical potential of the “dispositif” in different ways. In our case, the technical device is an assembly of several complex components:

- Splash <sup>(1)</sup>, software dedicated to real time videomapping on any surface;
- LivePose <sup>(2)</sup>, software for detecting human silhouettes in a real-time video stream;
- Mobilizing.js <sup>(3)</sup>, a library for prototyping interactive works and a platform for collective interactivity with mobile screens.
- SATIE <sup>(4)</sup>, a spatial audio engine, capable of handling audio displays of arbitrary architectures.

From these tools and their interconnection, the aim is to set up a generic technical “dispositif” that allows numerous artistic explorations from the material it makes manipulable: an immersive visual environment calculated in real time 3D, a set of data representing the bodies of the spectators in the experience’s space and, finally, a system of networking of the spectators’ smartphones allowing organising group interactions.

## Pipeline description

---

Our technical “dispositif” is composed of three software tools, but also of a set of scenographic and technical factors that we will describe here. The physical space of the experience is necessarily shared by the viewers because their co-presence is fundamental for us: like the exhibition of a non- technological artistic work, they are gathered in a dedicated place and form a group. This co-presence is a major starting point of our research-creation since we aim to question the relationship that the participants can maintain between them-selves through our “dispositif.”

Immersive video projection is the most prominent technical element that we use, as it can cover large surfaces, including the totality of the walls of the exhibition space. Whatever the architectural configuration of the space, a video projection system must be designed to transform the interior space into the equivalent of a 360° projection screen. This task is commonly called *video mapping* and, in our case, it is handled by Splash. 10 years in development, it allows segmenting images coming from any video source to broadcast them through video-projectors placed in the exhibition space. The physical placement of the video projectors is planned to cover the desired surface and will be virtually reproduced in Splash in order to adjust the parameters of the projected images deformation: it uses anamorphosis in order to give the viewer an illusion that they are all facing the same image made with a homogeneous perspective, a condition *sine qua non* of a feeling of visual immersion. This is called projection

mapping. Splash can use many different sources of image input ranging from HDMI acquisition cards, video/image files, or network transfers (i.e. NDI <sup>(5)</sup> feeds). A multitude of software can be used as image/video source, in our case, Mobilizing.js will provide the input.

Mobilizing.js is a research project consisting of an authoring software environment for artists and designers. The objective is dual: on the one hand, to promote the creation of interactive works on different forms of screen devices (computers, mobile screens, IoT, etc.) and, on the other hand, to be a platform for artistic creation dedicated to collective interactivity in co-presence. We have chosen Web technologies (Javascript in particular) as our technical building blocks because they are well spread and based on public standards. Two main sets make up the Mobilizing.js environment: a library and a platform. Based on a modular logic, the library of Mobilizing.js aims at gathering different functionalities in a coherent programming interface to make software art creation on mobile screens more accessible. As for the Mobilizing.js platform, its ambition is to propose a design and production environment for collective interactivity devices. The scripts created with the Mobilizing.js library, designed for a single user, can become the basis of a multi-user shared interaction "dispositif," using a local network to link mobile screens as well as computers together. The universality of the web browser, which can be run on desktops as well as on smartphones and tablets, is an important advantage here. Mobilizing.js integrates, among other things, a primitive real 3D rendering engine as well as the beginnings of an audio engine, allowing designing artistic creations compatible with Splash: one or several computers connected to the Mobilizing.js platform take care of rendering of images that will be transmitted to Splash, which will project them in the exhibition space in an immersive way. In parallel, the Mobilizing.js platform allows synchronising events (such as user interactions or automatically managed commands) between smartphones, which opens up the possibility of involving collectives of viewers through their interactions with their mobile screen, which in turn can allow control over what is represented in the video-mapping. In its internal architecture, the Mobilizing.js platform is based on Soundworks, a full-stack JavaScript framework for distributed WebAudio and multimedia applications developed by Ircam collaborators (<https://github.com/collective-soundworks>), which uses the WebSocket protocol to manage network communications.

This use of WebSocket allows the Mobilizing.js platform to be open to other environments so that they can share their data flows with it. It is this openness that allows LivePose to complete our device.

LivePose is a command-line tool which tracks people skeletons from a RGB or grayscale video feed (live or not), applies various filters on them (for detection, selection, improving the data, etc) and sends the results through the network (OSC and Websocket are currently supported). Live-Pose is able to do all of this in real time, processing live video streams and sending out results for each frame at 20-30 FPS. The data transmitted by LivePose is formatted in JSON, which is particularly compatible with Javascript and, therefore, with Mobilizing.js. The basic principle of LivePose is to recognize human silhouettes in an image and to deduce a skeletal structure composed of joint points and segments connecting them. It is therefore a skeletal model represented by a series of coordinates in the image space (2D) that forms the output of LivePose. Among the functionalities under development is the agglomeration of images captured by a cluster of cameras in order to reconstruct these skeletons in a coherent 3D space, which would allow us to represent these skeletons in 3D in Mobilizing.js and to use the information of positions in the space of the spectators in our works of collective interactivity.

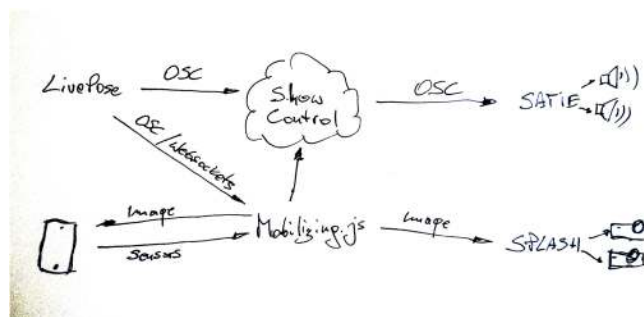


Figure 1: Our software pipeline data flow

The above-described software and hardware generic ecosystem is still technologically fragile and complex to implement. It lacks a GUI and robust communication interfaces. Yet, the software implements interoperability mechanisms which gives it considerable research-creation potential. But, in the infinity of the possibilities that artistic conception allows, it belongs to us to frame certain aesthetic and artistic orientations. We are, therefore, inspired by the technical elements of our own "dispositif" to choose our thematic directions. A first recent experiment between Mobilizing.js and LivePose was the occasion to lay the technical foundations of our "dispositif" while opening the tracks of artistic reflexion which we wish to explore in the future.

## Interactive installation example : Materia

Organized by Hexagram, a Canadian research-creation network in arts, culture and technology, in partnership with Elektra, an organisation that presents works and artists at the confluence of contemporary art and new technologies, *MATE-RIA: Public Laboratory for the Creation of Digital Knowledge* offered us an opportunity, in September 2022, to implement a first prototype of our device. Because it was an event dedicated to practice-based research in digital art creation and to the meeting between the actors of this community, we responded to the call for participation by proposing a research-creation workshop focusing on hands: how to link, through an artistic and technical “dispositif”, the bare hands of the participants, tracking movement and position with LivePose, and augmenting them with smartphones using their numerous embedded sensors?



Figure 2: Materia experimentation, photo A. Sermanson

Mobilizing.js supported the smartphone application, a large screen image display application, the networking of all the data flows in a dedicated local network, as well as a debugging application to check the status of the data transition in the network in real time. LivePose was modified specifically for this project to track users' hands as accurately as possible from a group of four cameras placed in the centre of a table. The goal was to reconstruct their positions in 3D in the space of the experiment. These data were transmitted in real time to Mobilizing.js, on the network dedicated to the experiment, which allowed producing graphical representations in 3D of the spectators' hands and thus, by extension, to define particular interactivities between these virtualized hands and the smartphones held by the participants. The hands positions and actions were, furthermore, sonified using SATIE. With the help of the workshop participants (*should we mention the names here?*), different scenarios of collective co-situated

interactivity based on the hands have been imagined. The result is the possibility of composing an ephemeral landscape using the traces produced by the movement of bare hands in the tracking zone of the cameras. The smartphone of each spectator is assigned a short sound, different for each one and conceived on the spot by the workshop participants. When the user touches the screen of the smartphone, it starts the recording of the movements of the hands that are present in its video capture area and the sound is played by the smartphone. SATIE sonification provides an ambient sonic texture linked to the detected presence of hands by LivePose. When the user releases the screen, the recording and sound playback stops. The traces left by the hands, visible in a dedicated screen that simulates an immersive environment, remain in the space for some time before being erased, which allows not to saturate the graphic space for too long and to leave room for new participants to the experience.



Figure 3: Materia experimentation, photo A. Sermanson

Thanks to this experience, which sketches a collective work of art, we were able to verify empirically that the artistic potential of our device was indeed considerable, as much as the technical difficulties of implementation that it reserves to us. It is important for us, at this stage, to properly situate our positions regarding the different components of this “dispositif”, especially since we were not able to implement projection mapping in this experiment. An artistic orientation nevertheless made its appearance implicitly: the power of contemporary computer vision systems, from which LivePose is built, and their implementation in public spaces leads us to a critical reflection on systems built for monitoring individuals. But before exploring this theme more deeply, it seems essential to us to get back to both the notion of immersion and the notion of collective interactivity.

## Sensory Immersion

---

Immersion technologies have been present in artistic practices for several decades. Their historical roots are well known and can be found in several in-depth studies (Olivier Grau - Virtual Art). It is accepted that the 19th-century panoramas are the origin of contemporary digital devices, as they share the same dream: to trick human vision in order to make the viewer believe that he is somewhere else than where he really is. As mentioned above, sensory immersion in an artistic device is achieved through the technical “dispositif” as a whole, including the physical space, which is augmented by dedicated elements, including audio and video projections. It is this adaptation of image projections onto a physical space that will contribute to the sensory immersion. The inclusion of the physical space in the definition of sensory immersion that we are making here implies a collective experience of immersion, common to all participants in the same space. We therefore exclude technical devices built around virtual reality headsets and other individual hardware, except in hybrid uses that maintain this fundamental relationship to physical space. The generic term for what immersive projection devices produce is spatially augmented reality. Once images can be projected onto large surfaces, several uses become possible. It can be a fully immersive space broadcasting views of a totally different place than the physical one, a visual augmentation of the physical place by scattered elements of decor or texture, or any balance between these two extremes, according to the continuum of mixed realities coined by Milgram et al. (Paul Milgram, Haruo Takemura, Akira Utsumi and Fumio Kishino, *Augmented Reality: A class of displays on the reality-virtuality continuum*, in *Proceedings of SPIE 2351, Telem manipulator and Telepresence Technologies*, 1995).

Video-mapping is one of the tools enabling this enhancement. It has many and diverse uses in scientific (planetariums), educational (Salame 2022, Puget 2019), museum (Nikolakopoulou et al. 2022), entertainment (Lee et al., 2015) and, of course, artistic (Lambert 2012 for example) fields. Its use to cover the periphery of publicly accessible physical space is the most common, and the technique for doing so is now well mastered: automatic geometric (Kurth et al. 2018) and colorimetric (Huang et al. 2017) calibration, standardisation of content representation formats, and dedicated hardware *off the shelf* show that it is now a known medium, although it still presents some complexity of implementation *in situ*.

As part of the many practices of video-mapping, spatial augmented reality is distinguished by the projection of graphic content into an interior space. The contents are interwoven with the physical space, which can become a medium of information, a window to a virtual world, or be transformed or even temporarily rendered invisible. The work *Displacements* (Michael Naimark, 1980, <sup>20</sup>) could be considered as one of the first applications of spatial augmented reality, clearly setting out the challenges of the hybridisation of realities, while the series *Bumpit* (Bertrand Planes, 2011)<sup>4</sup> is a more modern version of it and illustrates well the possibilities of transforming the perception of physical space within the framework of artistic installation. From a more technological and prospective point of view, the research project Room Alive <sup>18</sup> has demonstrated interactive uses of this spatial augmented reality involving several users: a room of daily life (e.g. a living room) is augmented with the help of an immersive image projection system with which the inhabitants can engage in ‘physical’ interaction, for example by walking on the images of small creatures evolving on the walls or on the floor. The co-presence necessarily induced by the spaces chosen for these experiments (living quarters) thus implied the development of a form of collective interactivity. On the basis of this observation, an experimental device produced within the framework of the same research project concerned remote communication between individuals, a form of video-conferencing of a new kind which proposes to video-map the interlocutors of a conversation in their respective spaces of physical presence [Room2Room: Enabling Life-Size Telepresence in a Projected Augmented-Reality Environment, <sup>22</sup>. These last references show us that interactivity plays an essential role in the perception of immersion (as shown by Hudson et al. 2019, interaction is as important when dealing with virtual objects as it is with real ones). But the implementation of collective interaction requires an appropriate technical device as well as a clear understanding of the very terms that make up this expression: collective interactivity. To clarify these terms conceptually is an imperative for us to conceptually establish the artistic “dispositifs” that our generic technical “dispositif” will support.

## Collective interactivity via smartphones

---

Collective co-located interactivity has been a new field of artistic experimentation since mobile screens (smartphones and tablets) became massively

widespread, in other words, at the turn of the 2010s, after the release of the Apple iPhone. Although the emergence of these devices can be considered as the starting point of a specific artistic research, working with mobility and the technical potential that it confers to these screens, the recent history of art and design presents some premises of collective interactivity.

Indeed, the artistic practices of interactivity have led to experiments concerning collective forms of digital interactivity, i.e. devices requiring the active participation of several people, often gathered in the same physical place. The colocation of the participants, turned to actors by the “dispositifs”, is one of the criteria of distinction with shared interactive systems and artworks, using the computer networks to create communities of geographically distant users, but joined together by the means of images, as in the cases of persistent virtual environments. “Localized media” are part of these practices that combine the mobility of devices equipped with positioning technologies (GPS) and the participation of a group of users. The artists’ studio Blast Theory, with experiences such as “Can you see me now?”<sup>5</sup> are among the pioneers of these practices.

Among the other fields of computer networks usages, we know that the state of the art of groupware, or multi-user interactive systems is very rich, in particular within Human-Computer Interaction (HCI). One of the many reference articles on the subject of *groupware* dating from 1991<sup>12</sup> gives a taxonomy of multi-user interactions on which current research in this field of computer science is still based.

The state of the art is much more limited regarding aesthetic “dispositifs” of situated collective interactivity. One of the few research projects that tend to combine interactive art making and locally networked creation with mobile devices is<sup>3</sup>. It consists in a creative environment using interconnected smartphones that allows users to play with a shared, locally broadcast musical environment. Made of a mix of technologies made compatible together using a data exchange protocol (OSC), each mobile becomes a kind of instrument whose individual manipulations are pulled into an audio composition broadcast over speakers.

A design-based research on collective interactivity from an HCI perspective helps to identify the major references in this field of research<sup>23</sup>. This paper aims to review a series of research works and interactive devices involving multiple people made by this research team based in Aarhus University. One of the assertions is that the whole set of arrangements must be taken into

account when designing collective interaction devices. This leads to the following concept of collective interaction: “CI focuses on how the interaction supports human-human interaction through the spatial organisation of people and coupling of the interaction”. Three main notions are considered as fundamental in the design of collective interactions: “Interaction Proxemics, Social Interaction and Co-experience.” Proxemics holds a particularly important role, since it conditions the type of social interactions that can be experienced by the collective of spectators/users. Indeed, the notion of proxemics comes from the social sciences’ research initiated by Edward T. Hall<sup>15</sup>. It can be considered as the science of behaviours between humans induced by the physical proximity of people in relation to each other, the organisation and structuring of space being able to influence these behaviours. A matrix based organisation is thus proposed to analyse the collective interaction devices. This analysis tool, as the authors specify, “should not be seen as depicting predictive, causal relationship, but rather as a way to illustrate how certain design strategies can be brought into play in order to pursue intended use qualities” (p.74). It is therefore a study that wishes to help in the design of collective interaction experiences. This kind of grid (or matrix) is also found in research carried out by our team<sup>19</sup> which also proposes an analysis of the constituent elements of a work using collective interactions (Group size, activity type, I/O Distribution, etc.) in order to extract a classification matrix.

The question of interaction with mobile devices, particularly smartphones, is very interesting on many levels that are out of scope of this article, but one of the areas that remain unexplored is measuring and understanding the *why* of some of the interactions in addition to *where* and *when*. This aspect is explored via a research topic and a corresponding application called MyExperience<sup>14</sup> which measures 140 parameters and follows the user around many different contexts of their everyday life. In addition, Dalsgaard et al.<sup>7</sup> offers a theoretical basis of the elements of *dynamic process of engagement* that an interactive experience can offer. Their analysis is based on three different types of installations in three different contexts: public art, museum installations and a department store interactive marketing medium. This is further expanded by Schroeter et al.<sup>24</sup> who position the *situated engagement* across three parameters: people, content and location, which have implicit cultural implications.

## Some foundations for a taxonomy of collective interactivity

## figures

---

One of the differences between these proposals and our position is that, in our research, it is not so much a question of collective interaction as of collective interactivity.

The ambiguity of the term interaction in the field of HCI is explored in <sup>16</sup>, which lists different concepts from which the HCI literature has based some meanings of the term (such as Dialogue, Transmission, Tool use, Experience or Control, summarized in a table p.5042). The term interactivity is also briefly mentioned (p.5041) as being both an equivalent and an alternative to interaction and, therefore, subject to discussion. The difference between interaction and interactivity is not discussed. However, the artistic practices of interactivity gives us a definition of the two terms as we have mentioned it earlier, about the twofolded "dispositif". We base our own definitions on the researches led by the team Aesthetics of New Medias (Paris 8 university) to define the notion of interactivity as being a relation which takes shape between an artistic "dispositif" and its addressees <sup>6</sup>. In the field of artistic creation, this relation is made possible by the capture of interactions engaged by the spectator when he use a physical interfaces and by a software structure that allows to "replay" this interaction. For example, the movements of the mouse is an interaction captured by the mouse, a physical technical device moved by the user's hand. This interaction is transmitted to the computer as numerical coordinates which, when used in a specifically designed program, allow the construction of different relationships with graphic or sound objects: moving a cursor around the screen, controlling the orientation of the point of view in a real-time 3D space (in video games, among others), or drawing colored segments in an area of the screen are all situations of interactivity enabled by the same interaction. Interactivity is thus built by programming out of interactions.

Research in HCI frequently build tools for analysis and help in the design of devices, because it needs to evaluate interaction systems quantitatively: the goal is to prove the effectiveness of one interaction technique in relation to another according to its context of use. Interactive art practice, on the other hand, seeks to establish non-quantifiable aesthetic relationships between the audience and the device that it encounters: it is a question of making an artwork in the strong and artistic sense of the term. This is the reason why, in our research, we propose a series of keywords that designate types of relations between people as starting points for the definition of collective interactivity

modalities and the conception of artistic devices implementing them. Our list is not exhaustive but currently consists of the following terms: collaboration, association, division, participation, cooperation, confrontation, inspection, supervision, competition, reconciliation or opposition. All of these terms qualify relationships between individuals and can be combined to formulate scenarios of collective interactivity works.

Our contribution therefore consists in the clarifying the foundations of a taxonomy of collective interactivities to be done. We should then elaborate, furthermore, about the possible entries of this taxonomy, define them in order to constitute some kind of guidelines to support the design of collective interactivities experiences in an already existing and operating technical context. This approach is therefore the opposite of the studies already carried out and quoted above.<sup>7</sup> But before presenting the first early draft of this taxonomy, we must complete the conceptual framework we are building here by addressing the theme of surveillance inferred by nowadays technologies, including the very ones that we use.

## Surveillance and control reversal

---

As we briefly mentioned previously, as LivePose is fundamentally a system for tracking individuals, it leads us to a consideration about population surveillance systems using computer vision technologies. This theme seems relevant to invest in our artistic works scenarios, knowing that to-day's "digital everything" (and mobile devices in first line) transforms our societies into people tracking and massive personal data storage machines, in public as well as private contexts. Contemporary art has already addressed this problematic reality of our societies. Rafael Lozano-Hemer,<sup>10</sup> David Rokeby and The Surveillance Camera Players are among artists who have addressed the questions of surveillance in different ways. As <sup>21</sup> says, the sense of being surveilled is highly dependent on an individual's place and era of birth, geographical location, and culture. By this, the artist means that some of us can be very sensitive about being watched by hidden devices because their daily living can be constrained by it, whereas others don't really care even if they know surveillance systems are lurking to spy on them. Artists try to show how these technologies can interfere with our existences and change the way we socially behave.



Using contemporary technologies to observe or record participants' actions in collective installations raises a lot of ethical questions. What kind of data is being recorded? How is data being used? Will the data be stored? Are there any identifiable features? Can we trust the security of the stored data? Some of the research listed above mentions these elements and takes them into account, such as in MyExperience application, which collects a lot of identifiable data. HCI research addresses these questions via standard legal devices. However there are known artistic works that allow themselves to be a commentary on some aspects of ethical questions. These questions are inherent to the potential of the technologies that are employed.

Accordingly, one of the widely explored themes in interactive, collective multimedia installations is (some form) of surveillance. The approaches vary greatly from reflections on the dangers of various forms of biometric tracking,<sup>8</sup> through generating empathy (Rafael Lozano-Hemmer, Level of Confidence (2015)), to playful critiques of public policies regarding implementations of different forms of biometric tracking,<sup>26</sup> to mention just a few. However, if we take into account all the possibilities of computer vision and tapping into mobile devices and social media to learn more about individuals, we realise how powerful such knowledge could become. One of the most prominent examples is China's Social Credit System (SCS) that quantifies social and civil integrity of all citizens and corporations,<sup>28,27</sup> and which is actually almost as old as China itself.<sup>17</sup> The most frightening part of China's SCS maybe its scale, as it is already used in several big cities of the nation. In such context, the data becomes not only a tool for interaction between humans and machines but also as a powerful weapon of control. The political, humanitarian and civil issues of such systems are out of scope for this article and are explored elsewhere<sup>27,25,9</sup> but we mention them here for a broader context.

We are conscious of the issues surrounding user data, privacy and consent when building multiuser interactive installations. We are presenting a prototype of an installation that uses many techniques for tracking users such as pose and action detection, position as well as interaction with a mobile phone and we propose an artistic and playful take on the subject of user consent to be tracked, as described, and how we can use that data in attempt to control the users, either individually or collectively. Such playful uses of topics related to surveillance are not new<sup>2,26</sup> but we believe that such approaches are necessary and efficient in raising awareness of these issues among some slices of populations.

The technological assemblage that we propose covers several aspects of a production pipeline that includes input (computer vision, interaction opportunity via smartphone), show control (timeline-based and interactive), and display (video mapping, audio spatialization). We aim to build a system and know-how that can be generic enough to be useful for other types of artistic, design, and social manifestations. Our prototypical installation offers only one of the possible views and uses. We hope that our solution can serve as a starting point for others.

- 
- (1) <https://sat-mtl.gitlab.io/documentation/splash>
  - (2) <https://sat-mtl.gitlab.io/documentation/livepose>
  - (3) <https://mobilizing-js.net>
  - (4) <https://sat-mtl.gitlab.io/documentation/satie>
  - (5) video-over-IP transport and codec

## Conclusion: Draft of a taxonomy by example

---

We will conclude with three hypotheses of artistic installations based on the conceptual positions and technical choices we have outlined throughout this article. This work relies on the generic "dispositif" that we are setting up: an immersive "big picture" realized by video-mapping, a tracking system of the spectators in the space and smartphones interconnected with a local network, the whole activated by a group of spectators invited to share this experience. Surveillance is the general theme that we choose for these experiments, which allows us to contextualize our subject and to lead our choices of implementations. In order to ensure a certain critical and reflexive dimension, we aim to provide a kind of situation reversal in the experiments, the controlling power passing from one side to the other: in a first time, the spectators interact with a system which relies on informations captured about them (their position in the room, the proximity between them, the gestures they produce with their smartphone, etc.) and, in a second time, this system reveals itself to everyone by showing the "other side of the story", i.e., the information that has been used by the system to function. The black box becomes transparent. The visual elements that will be used will be mainly textual: words, sentences, paragraphs will be displayed in the smartphones, as in the immersive projection. The relation between the spectators and these textual

elements will be regulated by the “collective interactivity figures” that we retain for these first effective artworks : collaboration, cooperation and competition.

## Collaboration

---

The immersive projection shows a series of white-on-black text excerpts about surveillance and about textual content control systems in different contemporary contexts. The scale of the ethical “seriousness” of these systems would go to a crescendo. One can think of the erroneous corrections produced by automatic spellcheckers when writing SMS, to the censorship and generalized surveillance exercised on social networks in certain countries with totalitarian political regimes that can lead to the incarceration of militants, for example. These texts do not appear complete, some words are missing. The audience has to fill in the gaps in these texts with their smartphones. The texts appear in a sequential way: a first text is displayed, when it is completed by the public, it gives room to a second text, and so on. This principle allows the progression in the “seriousness” of the following texts.

In their smartphones, the spectators see a circular list of words displayed that they can rotate to select one of the words. A “send” button, similar to an instant messaging interface, allows the word to be transmitted to the text in the immersive image. This list of words changes according to the proximity of the people with one of the empty spaces in the text: it is necessary to be close to one of the holes of the text to be able to fill it, or else, the screen indicates “out of reach”, meaning the impossibility of participating. Thus, if the spectators do not collaborate together by spreading out in front of the text, they will not be able to fill it in to discover the next text.

The last text that appears is a report of the interactions that took place during the experiment, a description exposing the data collected by our “dispositif” to make the artwork operate: the number of participants over time and, for each one, their number of right and wrong answers, the time needed to complete the texts, the number of participants connected but who did not collaborate (being all the time “out of reach”), the models of smartphones used, etc. A form of natural language visualization of the data collected throughout the experience, reminiscent of “bots” or automatic conversational agents.

## Cooperation

---

Cooperation and collaboration are distinguished by the temporality of group interactions. Collaboration does not necessarily imply a temporal concurrency, the actions complete each other even if they are not synchronous, as in the example above. Cooperation implies, according to us, at least a proximity, and to the most a temporal synchronization between the interactions, the actions must be operated jointly by all or part of the group of spectators.

The immersive image has a white, plain background. Each spectator sees a white word on a black background appearing in his smartphone. When they tap the screen of their smartphone with their finger, the word appears in the immersive image in the place where it belongs in a sentence. If all the spectators tap their screen with their finger, all the words appear in the projection, but a certain order must be respected in these interactions, otherwise the words disappear just after their appearance in a fading effect: the spectator who sees the first word of the sentence in his smartphone must be the first to tap his screen, followed by the second, then the third, and so on until all the spectators have made their word appear in the large image in the right order of the sentence. When a sentence is completely displayed, the colors are reversed in the screens, the white background becomes black and the text is drawn in white and vice versa on the smartphones.

This sequential collaboration between the spectators will allow them to progressively discover sentences coming from texts about industrial espionage cases based on security flaws in computer systems.

## Competition

---

The immersive image is empty of text and is split in two coloured areas: white on the left and black on the right. In their smartphones, viewers are asked to write their nickname before they can see an editable text field (a prompt) centred in their screen. Depending on the person, the background colour of the screen is black or white. The group of spectators is actually split in two using this colour: if there are 10 spectators, 5 will have a black background and the others will have a white background. Under the prompt, on the mobile screen, the following inscription can be read: “Look for those who do not have the same background colour as you and report their misdeeds!” Using the prompt, viewers

can write what they want to report. An automatic moderation system checks whether the sentence is a denunciation or not. If they are successfully recognized as denunciations, the sentences are written in negative in the corresponding coloured area in the immersive image (i.e. white text on a black background). If not, the viewer's screen is marked with an error message and the prompt is cleared so that he or she can try again.

If no one writes with their smartphone after a certain period of time (30 seconds to 2 minutes, to be verified in real conditions), sentences from a collection made beforehand and integrated into the program are displayed from time to time in one of the coloured zones of the immersive image. The idea is to introduce doubt: did someone write this, or is the device expressing itself?

The group that manages to fill its screen with sentences first wins the right to write whatever they want without any automatic moderation and on the whole immersive image, which will take on the background colour of the winning group, for 1 minute. After that, a list of all the sentences written during the whole session with the associated user's nickname are displayed for a few seconds, exposing everyone's actions to the eyes of the public, before the device resets.

These 3 scenarios are still to be refined in their details, but they show how a key notion can be used as the root of an artistic proposition using our "dispositif". Our future work consists in concretely realizing these scenarios and putting them to the public test.

## Acknowledgements

---

We would like to thank our Materia workshop participants: Annaëlle Sermanson, Idun Isdrake, Heather Kelley, Valentine Auphan, for their involvement along the path, Hexagram and Elektra for facilitating the workshops and contacts with other creative humans.

## References

- 1 Giorgio Agamben, *Qu'est-Ce Qu'un Dispositif?* Rivages poche, Petite Bibliothèque, 2007.
- 2 Anders Albrechtslund, Lynsey Dubbeld, "The Plays and Arts of Surveillance: Studying Surveillance as Entertainment", In: *Surveillance & Society* 3.2/3, Sept 2002, ISSN: 1477-7487, DOI: 10.24908/ss.v3i2/3.3502.
- 3 Jesse Allison, Yemin Oh, Benjamin Taylor, "NEXUS: Collaborative Performance for the Masses, Handling Instrument Interface Distribution through the Web", In: (), 6.

4 *BumpIt! Main Page*, <http://www.bertrandplanes.com/bumpit/>.

5 *Can You See Me Now? — Location Based Game*, <https://www.blasttheory.co.uk/projects/can-you-see-me-now/>.

6 Jean-Marie Dallet, "Figures de l'interactivité", In: *Anomalie*, Mar. 2003.

7 Peter Dalsgaard, Christian Dindler, Kim Halskov, "Understanding the Dynamics of Engaging Interaction in Public Spaces", In: *Human-Computer Interaction – INTERACT 2011*, Ed. by Pedro Campos et al. Vol. 6947, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, 212–229, ISBN: 978-3-642-23770-6 978-3-642-23771-3, DOI: 10.1007/978-3-642-23771-3 17.

8 David Rokeby, *Border Patrol (Collaboration with Paul Garrin)*, <http://www.davidrokeby.com/border-patrol.html>.

9 Ann Dennon, *Students on School-Issued Devices Are Under Constant Surveillance — BestColleges*, <https://tinyurl.com/4sastf6u>, oct 2021.

10 Brian Droitcour, *Rafael Lozano-Hemmer on Seductive Participation and the Oppression of Metrics*, June 2020.

11 Anne-Marie Duguet, "Dispositifs". In: *Communications* 48.1, 1988, p.221–242, DOI: 10.3406/comm1988.1728.

12 Clarence A. Ellis, Simon J. Gibbs, Gail Rein, "Groupware: Some Issues and Experiences", In: *Communications of the ACM* 34.1, Jan. 1991, 39–58, ISSN: 0001-0782, 1557-7317, DOI: 10.1145/99977.99987.

13 Michel Foucault, *Dits et Écrits*, Vol. III, 1977.

14 Jon Froehlich et al. "MyExperience: A System for *in Situ* Tracing and Capturing of User Feedback on Mobile Phones", In: *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services - MobiSys '07*. San Juan, Puerto Rico: ACM Press, 2007, p.57, ISBN: 978-1-59593-614-1. DOI: 10.1145/1247660.1247670.

15 E Hall, *The Hidden Dimension*, Anchor, 1966.

16 Kasper Hornbæk, Antti Oulasvirta, "What Is Interaction?" In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2017, 5040–5052, ISBN: 978-1-4503-4655-9, DOI: 10.1145/3025453.3025765.

17 Min Jiang, "A Brief Prehistory of China's Social Credit System", In: *Communication and the Public* (), p.6.

18 Brett Jones et al. "RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units", In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, Honolulu Hawaii USA: ACM, Oct. 2014, 637–644, ISBN: 978-1-4503-3069-5, DOI: 10.1145/2642918.2647383.

19 Oussama Mubarak, "A Taxonomy of Multi-User Co-Located Interaction Environments", In: (), p. 4.

20 Michael Naimark, *Displacements*, <http://www.naimark.net/projects/displacements.html>.

21 B.D. Owens, "Watching Me, Watching You: Reflections Upon Surveillance, Gare Loch Duality and the #UndesiredLine", In: *Media-N* 17.2 (Oct. 2021), 49–68, ISSN: 1942-017X, 2159-6891, DOI: 10.21900/j.median.v17i2.882.

22 Tomislav Pejisa et al. "Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment", In: *Proceedings of the 19th ACM Conference on Computer-*

Supported Cooperative Work & Social Computing, San Francisco California USA: ACM, Feb. 2016, 1716–1725, ISBN: 978-1-4503-3592-8, DOI: 10.1145/2818048.2819965.

23 Marianne Graves Peterson, Majken Kirkgaard Rasmuseen, Peter Gall Krogh, “Collective Interaction: A Designerly Visual Analysis of Seven Research Prototypes”, In: Proceedings of the 29th Australian Conference on Computer-Human Interaction, Brisbane Queensland Australia, ACM, Nov. 2017, 68– 76, ISBN: 978-1-4503-5379-3, DOI: 10.1145/3152771.3152779.

24 Ronald Schroeter, Marcus Foth, Christine Satchell, “People, Content, Location: Sweet Spotting Urban Screens for Situated Engagement”, In: Proceedings of the Designing Interactive Systems Conference on - DIS '12, Newcastle Upon Tyne, United Kingdom: ACM Press, 2012, 146, ISBN: 978-1-4503-1210-3, DOI: 10.1145/2317956.2317980.

25 Saif Shahin, Pei Zheng, “Big Data and the Illusion of Choice: Comparing the Evolution of India’s Aadhaar and China’s Social Credit System as Technosocial Discourses”, In: *Social Science Computer Review* 38.1, Feb. 2020, 25–41, ISSN: 0894-4393, 1552-8286, DOI: 10.1177/0894439318789343.

26 Karen Louise Smith et al. “Playing with Surveillance: The Design of a Mock RFID-based Identification Infrastructure for Public Engagement”, In: *Surveillance & Society* 9.1/2, Nov. 2011, 149–166, ISSN: 1477- 7487, DOI: 10.24908/ss.v9i1/2.4108.

27 Alexander Trauth-Goik, Chuncheng Liu, “Black or Fifty Shades of Grey? The Power and Limits of the Social Credit Blacklist System in China”, In: *Journal of Contemporary China*, Sept. 2022, 1–17, ISSN: 1067-0564, 1469-9400, DOI: 10.1080/10670564.2022.2128638.

28 Karen Li Xan Wong, Amy Shields Dobson, “We’re Just Data: Exploring China’s Social Credit System in Relation to Digital Platform Ratings Cultures in West- ernised Democracies”, In: *Global Media and China* 4.2, June 2019, 220–232, ISSN: 2059-4364, 2059- 4372, DOI: 10.1177/2059436419856090.