

Hello, An Interactive Cinematic Generative Artwork

Pedro Alves da Veiga

Centro de Investigação em Artes e Comunicação
Universidade Aberta, Universidade do Algarve
Portugal
me@pedroveiga.com

Abstract

“Hello” is an interactive cinematic generative artwork that gained traction within the Cultural Adventures research project, an international endeavor between Portugal and Thailand. The encompassing rationale assumed the need of a link connecting both countries, and thus “Hello” was born from the similarities that can be found in commuters that flock to and from both capital cities by train, often more engaged with their mobile phones than with their immediate surroundings – people, train or traversed landscapes –, which constitute more of a shapeless blurred audiovisual background than a sharp attention-grabbing focus. The artwork is also used as a demonstrator of the potential of generative art in producing complex cinematic narrative figurative artworks, yet permeated with interference, chaos and randomness, and how the audience perceives and makes sense of them through the ability to identify and create patterns, to conceptualize, to deal with information overload, and to imagine and build a narrative through abductive reasoning from incomplete yet suggestive information.

Introduction

Common human everyday visual perception is pragmatic, and is mostly oriented toward the identification of objects and shapes in visual scenes. Individuals report perceiving objects in pictorial compositions even when those compositions are devoid of recognizable objects (Ishai, Fairhall and Pepperell 2007), which points to an inescapable human trait: that of searching for patterns.

Inspired by the work of Ezra Pound, McDonald (1993) proposed that by opening up the perceptual field to the indeterminacies of fragment and chaos, it became possible to recreate a multi-sensory field within which audience and creator encounter the same stimuli toward a desired end. This process enables the role of quester (after patterns, knowledge or meaning) to be held simultaneously or passed around audience members, who need to chart an unknown, unstructured territory, collecting partial findings as the exploration takes place in real time. The absence of (expected) markers unsettles narrative assumptions and challenges the audience from the comforts of convention and its passive position into an experience of navigation by

periplus. The audience is no longer outside the artwork and its narrative, but inside it, mapping its shape and meaning as they gather information and act upon it.

The term *generative* implies an algorithmic framework within which takes place the creation of the artwork itself. The algorithm combines structure with controlled randomness, bringing order and chaos together, in line with Pond’s multi-sensory field, and each iteration becomes the seed for the next iteration, thus resulting in a seemingly infinite, non-repetitive variety of states within a certain aesthetic boundary defined by the artist / programmer (Galanter 2014).

“Hello” is an interactive cinematic generative installation that outputs a constant stream of images and sound, which can, to some extent, be influenced by the audience. The installation foresees two different audience behaviors as it is regarded both as a chaos generator and as a system exciter. If no – or little – movement is detected in the exhibition room, the (still) audience is presented with a continuous stream of a flickering mosaic set to a background track of ambient train-related noises. All the images used to produce this stream are of Thai people, landscapes, trains, platforms and stations. All the sounds were captured in Portuguese trains and train stations. This sets the mood to an imaginary journey, where images are deconstructed and reconstructed, much like what happens in a commuter’s mind, traversed by the fleeting landscapes, glimpses of their fellow travellers and of the their own imagination, triggered by thoughts, conversations and other stimuli. However, a curious and more dynamic audience will be presented with a stream of short text messages, half of which are in Thai (left side of the projection) and the other half in Portuguese (right side of the projection), visually inspired by most chat and messaging systems currently in use and replicating their expected behavior of moving upward as new utterances are introduced into the dialogue. This dialogue becomes more active as the artwork detects more movement in the exhibition room, which also impacts the several parameters of the mosaic generation and sound intensity, as the installation echoes the action in the exhibition room. An added element of surprise is brought to the equation as the system captures random video frames of the audience and introduces them among the

original Thai imagery, literally bringing the audience into the artwork.

Figure 1 shows a run-time screen-capture of “Hello”, with Thai (left) and Portuguese (right) dialogues underway.

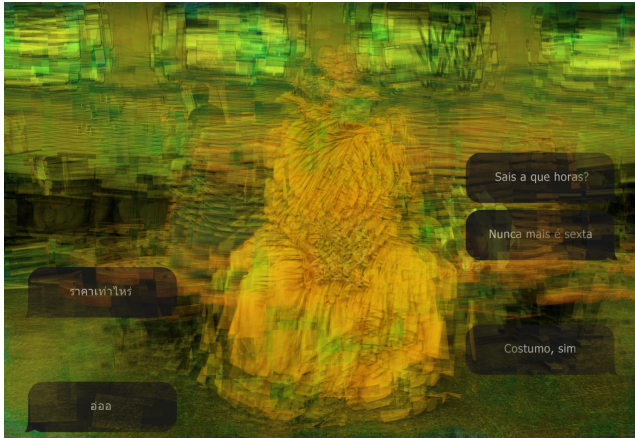


Figure 1 - Screen capture of "Hello", with Thai and Portuguese dialogues underway, over a cinematic background.

The dialogues in “Hello” contain hints of a love-crime fiction, spanned and scattered across multiple utterances from both interlocutors. As the audience realizes that the more movement is detected in the exhibition room, the more utterances are provided, thus allowing them to access more information and complete the narrative, it is expected that the exhibition room then may become the stage of an audience performance, both physical and intellectual, as members of the audience may engage in collaborative speculation and debate.

A video with a real-time capture of "Hello" is available at <https://vimeo.com/alvesdaveiga/hello>.

Methodology

Due to its nature, this project – both the artwork development and the related research – were brought to term using *a/r/cography* (Veiga 2019), an arts-based research methodology derived from *a/r/tography* (Springgay, Irwin and Kind 2005). This methodology stipulates seven generative, mutually influencing steps and its outcome involves the artwork, the documented research, the communication of any project related outcomes, including public exhibition and presentations, and all the inherent processes. These steps are presented over the next sections.

Inspiration

The concepts of communication and transportation have some interesting similarities and intersections: trains connect distant locations and mobile devices connect distant people; trains carry our physical presence, while mobile devices carry our immaterial presence. But a train track also divides territories, in the same way that a mobile device can separate people who share a common location, by allowing them to easily disconnect from their physical surroundings and engage with a remote (non-local) physical

or virtual technology-mediated reality. This was rendered evident through observations in public transportation in the greater Lisbon area, in Portugal, where local commuters are often absorbed in the use of their devices rather than in conversations with nearby fellow travellers or showing an interest in the visual details around them, both inside and outside the train.

Trigger

Similarities in this type of behavior were highlighted during conversations with Phanuphong Songkhong, a Thai designer, who documented very similar observations in several thousand drawings he produced during his own commutes, in and out of Bangkok.

Acknowledging so many similarities in distant countries, with such different cultures, was the trigger that set in motion the whole project.

Intention

From the original inspiration stemmed the will to translate into an artwork how the whole sensorial stimulation during such train journeys provided a mosaic of mixed sensations in the brain, some of which were the direct result of vision and hearing, others the mental offspring of the activities people engaged in (texting, e-mailing or telephone conversations), and the rapidly changing mosaic-like frame of mind that ensued.

There was also the intent to portray a reality shared by Portuguese and Thai daily commuters, where technology mediated dialogues were apparently more easily established than local dialogues, as the majority of regular commuters were usually more deeply engaged in interactions with their mobile devices, rather than having conversations with their travel companions.

All these factors contributed to the intention of delivering a mosaic-like atmosphere, where stimuli are segmented and blurred into an all-encircling and all-absorbing yet challenging experience, without requiring the use of virtual or augmented reality to accomplish its immersive effect – much like the situations it depicts.

Conceptualization

All the overlapping, mosaic-like stimuli found in a train commute contribute to creating a sensory journey of fleeting visions and sounds, driven by our immediate physical surroundings, by conversation-triggered memories or imagined scenes, or even our own personal ramblings and self-awareness.

Replicating this sensory stream will invite the audience to embark on a cross-continental immersive and interpretive journey. As visitors enter and move in the exhibition room, they are already and unknowingly interacting with the installation, potentially becoming “the subject of conversation”, as images of the room are captured and inserted into the visual stream, and the “journey conductor”, because their movements in the room directly affect system behavior.

As the audience is presented with a dialogue, of which (probably) only half is intelligible, they are also invited to imagine and reconstruct a narrative through abductive reasoning from incomplete yet suggestive information (Peirce 2012).

Related work

Generative Cinema may not be a widespread genre but has a long history rooted in earlier forms, among which is Calculated Cinema. For Bonet (2007) the Calculated Cinema series intends to open a window of observation allowing a glance both backwards in time, as well as forwards, offering "an antidote for the digital animation styles which are now hegemonic and omnipresent" (Bonet 2007).

In Calculated Cinema, directors such as Sergei Eisenstein and Dziga Vertov laid the foundations of metric or rhythmic montage, adopted in the current artwork.

Manovich and Kratky (2005) consider this approach to be at the heart of Soft Cinema: if *hard* cinema is based upon storylines, scripting, shooting and editing guidelines, Soft Cinema is based upon algorithmic sequences of visual units extracted from a previously constructed database, often including randomness in their selection mechanisms. Soft Cinema advocates the replacement of single-screen images by mosaics of images. Each new screening will differ, due to the randomization of the database units and large number of possible combinations.

Soft Cinema and Calculated Cinema are thus both at the heart of Generative Cinema, which specialises in the delivery modes of a heterogeneous realm of artistic outcomes based upon the combination of predefined elements (order) with different factors of unpredictability (chaos) in its conceptualization, production and presentation (Grba 2017, 8).

In "Hello" generative processes become richer and more complex through the introduction of probabilistic and random factors in predetermined rules, thus producing successive different generations of audio-visual output where each newer generation inherits structural and compositional characteristics of the previous one(s), as well as randomness and non-linearity.

Prototyping

The idea to portray such a multisensory mosaic took shape and the project was founded around three key areas: movement detection, dialogue generation and a cinematic effect obtained from still images.

Movement detection

The interaction is built upon the analysis of consecutive frames, captured by a webcam facing the audience. This analysis takes into consideration noise levels – higher noise levels in darker environments, lower levels in well-lit environments – and a pixel-by-pixel image differential (Singla 2014; Suresh and Lavanya 2014):

$$\Delta I_{(i,j)} = ICurr_{(i,j)} - IPrev_{(i,j)}$$

This allows for the determination of the general area of the room where most movement occurs, i.e. where most

different pixels between consecutive frames are found. Since the purpose of movement detection in this artwork is merely to determine the prevalence of utterances on the Portuguese side (right) or the Thai side (left) of the projection, its calculation is simple enough to be undertaken in real-time with no hindrances to the overall artwork performativity.

Dialogue generation

The dialogue is generated from a database of utterances, some of which – Portuguese – were obtained from the author's own SMS conversations, while others – Thai – were provided by Thai students from King Mongkut's University of Technology Thonburi, in Thailand. These utterances were later complemented by specially created content in both languages, in order to weave the underlying plot.

Since the text-based dialogue is presented in those two seldom-learned together languages, it is fair to expect that no audience member will know both of them, and here lies the heart of the game: the purpose of any dialogue is to exchange information between two interlocutors. When presented with only one intelligible information source, the audience will gather key segments of information about a certain topic from different utterances, some of which are meaningful (e.g.: "He robbed her?", "She fooled everyone", "It was a red house") while others are conversation fillers (e.g.: "I see", "Is that so?") leading the audience into believing that the key information was on the other (unintelligible) part of the dialogue. A total number of 400 utterances were produced in both languages.

The presentation of different utterances follows the established *chat-room* paradigm that determines newer messages should appear at the bottom of the screen and push older messages upwards. The system was thus implemented with a FIFO message queue for each language, storing the screen coordinates for every utterance. As the screen coordinates are updated, if one utterance exits the screen (negative y coordinate) then it is deleted from the queue. Each utterance has a time-to-live (TTL) property, which determines that older messages will be discarded after their life-span reaches zero. The purpose of this mechanism is to clear the projection space of all messages if no movement is detected for a certain amount of time – corresponding to a group of visitors having exited the room or being very still for very long. If the queue is empty, then the first message that will next be presented in either language is "Hello" (Olá, สวัสดี).

Each time an utterance is displayed, it is marked as used (Boolean state), to avoid repetition of previously and recently used utterances. After all utterances have been used, they are marked as unused and the process restarts.

A compensation mechanism was also implemented to prevent messages in Portuguese or Thai alone from being consecutively generated, disrupting the feel of a dialogue. Through experimentation it was determined that after a maximum of three consecutive messages in either language, even if movement is still being detected on the respective side of the room, a message in the other language

will be displayed, thus adding to the likelihood of the dialogue.

Cinematic Effect

Here resides the aesthetic core of the artwork, as image fragmentation and subsequent reconstruction processes define its visual identity.

Three very different creative minds and works inspired the algorithm behind the main animation: Marshall McLuhan, Ezra Pond and David Hockney. McLuhan (1988) claimed that the great technical possibilities of the cinematograph involved the “perceptions of simultaneities”. For him, this was achieved through a mosaic, as a world of intervals, not very dissimilar to what David Hockney suggests through his collages, criticizing the shortcomings of traditional photography, which limits the observer to a single, frozen perspective. He likens this frozen view to that of a paralyzed Cyclops (Vrobel 2011). Pond further unifies these two visions by stating that an image presents an intellectual and emotional complex in an instant of time. For him the image is not an idea, but rather a radiant node or cluster. He proceeds into calling it a *vortex*, “from which, and through which, and into which, ideas are constantly rushing” (Pond 1914).

“Hello” uses two different image sources: a set of 148 photographic images, categorized as either “people” or “landscapes”, and the webcam stream. Since “Hello” was developed in Processing, it takes advantage of the draw() function as a continuous loop through which different generations are produced. At any given moment one image is chosen as the main image. This main image is fragmented into a mosaic of sub-images, and each sub-image is modulated in terms of scale, rotation, tint and transparency:

$$I_m = \sum_m s_i(x, y, fc, \mu, s_{i-1}).r_i(x, y, fc, \mu, r_{i-1}).tr_i(x, y, fc, \mu, tr_{i-1}).ti_i(fc, \mu)$$

Scale (s), rotation (r) and transparency (tr) modulation are complex procedures, combining trigonometry and controlled randomness as functions of spatial position, frame count, audience interaction (amount of detected movement) and previous values (previous generation). Tint (ti) is only modulated as a function of frame count and detected movement. The end result produces a sensation of flickering movement, as if the (still) images are in fact originally imbued with movement, as the mosaic evolves over time. The same modulation procedure is applied to an “interference” image, but only a small amount of its sub-images are displayed over the main image, as shown in figure 2. These interferences are subtle, yet quite visible when animated and they constitute an interpretation of both McLuhan’s perception of simultaneities and Pond’s vortex.

The consistent and subtle progression of a dominant hue and the use of transparency contribute to easing each new main image mosaic over its predecessor, thus providing the sense of a continuum, where new images take some time to fully develop and be perceived by the audience, documented by figure 3.

The webcam stream is used sparingly, very seldom retaining one of the captured frames (< 5%) for movement

analysis and using it either as a main image (as depicted by figure 4) or an interference image. As they are interwoven in the animation flow, the audience will usually take some time before they fully process and perceive them, acting as an added element of surprise and bringing the public into the artwork itself.



Figure 2 – The white circles show the interference sub-images over the main image.

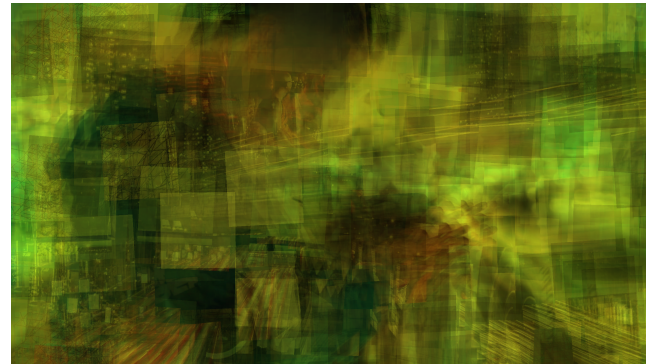


Figure 3 – The merging of two consecutive main images as the dominant hue is shifting from green to yellow.

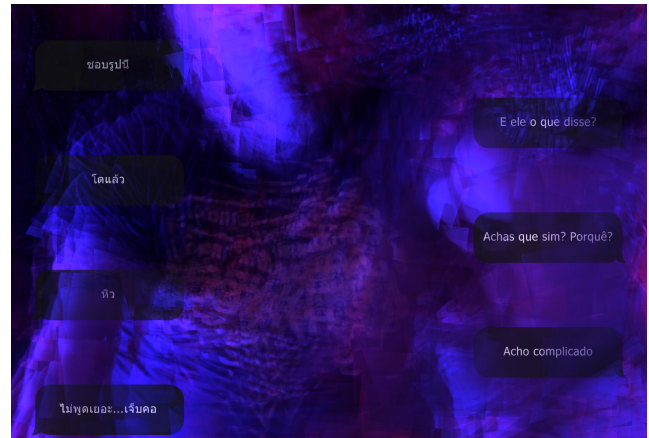


Figure 4 – A glimpse of the author, captured from the webcam stream.

The audio in “Hello” was captured in Portuguese trains and train stations. These recordings incidentally captured fragments of actual ongoing conversations and announcements during suburban commutes in the Lisbon area,

where trains are not only filled with Portuguese citizens, but also a very significant number of tourists, thus turning the soundtrack into a rather intricate mesh of different languages. These recordings were creatively mixed into four loops, corresponding to four floors of system excitement.

The excitement state (ES) is calculated into a variable as a function of the detected movement. This variable is controlled by a TTL mechanism based upon a weighted arithmetic mean, allowing for the soft transitioning between very different amounts of movement detection, fading in and out. The first loop corresponds to ES-0 (quiet room), is the default soundtrack and spans over two minutes and thirty seconds, looping throughout the whole exhibition.

The three remaining loops correspond to ES-1, ES-2 and ES-3, and were produced so that their playback can start and end seamlessly. They are consecutively triggered as the system reaches each of the three remaining excitement states. No new playback is started for a particular ES if the corresponding sound is already playing. The soundtrack thus mimics the events in the exhibition room, becoming very loud and frantic when ES-3 is reached, corresponding to a very agitated audience. As the movement quiets down so does the soundtrack, as each file playback (ES-1 to ES-3) reaches the end.

Hello was entirely developed in Processing 3.3.7. Its graphic vocabulary consists of image objects alone (PImage) and the soundtrack is delivered through overlapping looped samples manipulated with the Minim sound library.

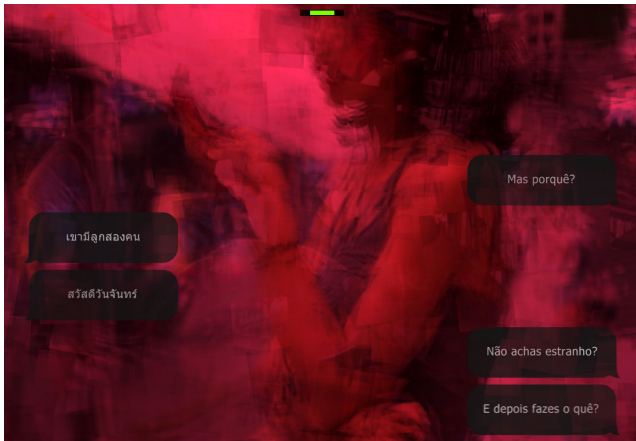


Figure 5 – Calibrating system sensitivity with the help of a color-coded area of a small viewport, top center of the image.

Testing

Implications

The artwork was due for simultaneous exhibition in two venues separated by over 10,600 km. This posed another challenge since, as stated in the Movement Detection discussion, webcam noise needs to be taken into consideration in order to effectively distinguish between noise (mostly caused by poor lighting) and actual movement (caused by visitors), which means that the installation needs to be calibrated locally.

These simultaneous events would prevent the author from being at both locations for the opening dates, thus the solution was to integrate interactive calibration into the artwork, taking into account the noise floor and movement detection threshold, including ES differentials. To this effect a small viewport depicting the ES can be rendered visible at all times, and direct testing under the actual exhibition room conditions allows to accordingly control the system sensitivity, as shown in figure 5.



Figure 6 – Poster for the collective exhibition held in Faro and Bangkok.

Intervention

“Hello” was simultaneously presented at Clube Fareense, in Faro, Portugal and at the Humble Projects gallery, in Bangkok, Thailand, as part of the “Conversa” (conversation) collective exhibition (figure 6).

The deployment at each venue was made according to the rider shown on figure 7, reproducing – to some extent – a cinema environment, while allowing for the free flow of visitors in the screening room.

Audience feedback

During the attendance of the Faro exhibition opening, the author was able to interact with several visitors, and assess their opinions and reactions through informal interviews.

This choice of format is justified by its openness to incorporate different visions and insights, which may have been absent from the author’s intentions, thus allowing for a better exploration of audience perception.

Based on 15 informal interviews, the author was able to ascertain that:

- 5 Portuguese visitors correctly inferred that the first message in Thai was “Hello”, even though they had never seen the word before and admitted to being incapable of copying the writing. They developed some curiosity about the language and posed a number of questions about it;
- 8 visitors failed to establish a causal connection between their movements and the artwork’s behaviour, especially at times when there were too many people in the room. The correlation was more easily established when only one individual or smaller groups were present, because as soon as they walked into the room and their movement was detected, they were greeted by the “hello”

messages. This contributed to establishing a sense of cause and effect.

- 3 individuals, who came in a group, were intrigued by the dialogue flow and were discussing whether it was purely random (which it is) or if there was also a link to their actions. This added to their performance since they decided to explore different situations (sitting down, getting on their knees, jumping, etc.);

- a group of 5 visitors interacted the longest with the installation (3 minutes, over six times the mean amount of time spent, established by Smith, Smith and Tinio (2017)), and were actively collecting information from the dialogue and discussing it in order to understand the underlying plot. When asked about what triggered their curiosity they stated it was the message about “someone being robbed”;

- 1 individual stated that, when alone in the exhibition room, he purposefully placed himself out of the webcam’s range, thus avoiding the text-messaging aspects of the installation, just to appreciate the cinematic experience, which he described as “mesmerizing”.

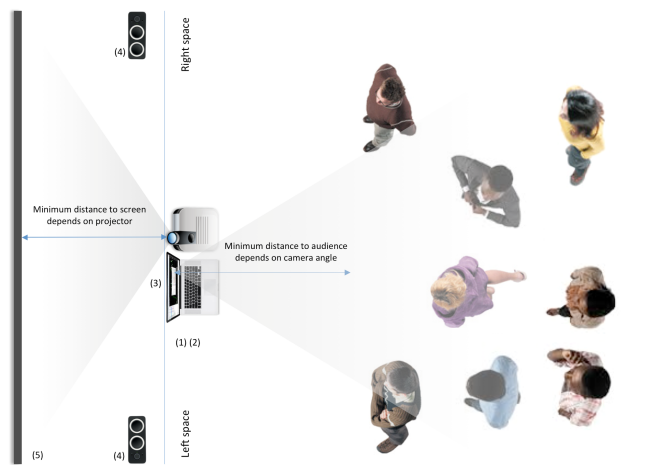


Figure 7 – The installation rider: (1) laptop or desktop computer (Windows 8.1 or above, Processing 3.*, OpenCV Library, Video Library, Minim Library, HDMI or VGA output; (2) webcam; (3) high-resolution short throw video projector; (4) stereo speakers; (5) wall or projection screen.

- all 15 individuals confirmed the overall sensation of being in a train or in a train station and two of them even vented the possibility that the conversation (which was also the name of the exhibition) might actually be taking place between Faro and Bangkok in real time.

Conclusion

The present article is the first formal written intervention within the *a/r/cographical* project. It depicts and presents all seven stages of its creative development, from inspiration to intervention, as well as the findings that were gathered during the attendance of the Faro exhibition opening.

The relationship the artwork established with its interactors managed to create both an individual and group performative and interpretive experience. However, smaller group interactions seemed to be the most successful in establishing the interaction correlations, particularly in collaborative exploration of the text-messages and finding meaning in the artwork.

As the curiosity to unravel the underlying story increased, so did the audience restlessness, often disregarding any initial constraints or awkwardness related to self-awareness. Speculation and interaction with other audience members then became almost natural, in a collaborative effort to decipher the plot. This suggests that some of the utterances, in future versions, may become more impactful (quizzical, strange or even shocking), in order to better capture the audience's attention.

Departing from very incomplete information some audience members were able to identify the key aspects of the underlying narrative, based upon key utterances, and regardless of their order.

The narrative was thus built fragment by fragment, as Ezra Pond suggested in his *ideogrammic* method, the culmination of the *periplum*, as the method of “presenting one facet and then another until at some point one gets off the dead and desensitized surface of the reader’s [audience’s] mind, onto a part that will register” (Pond 1970).

These conclusions are in line with Dewey’s statement (1980) that “the product of art – temple, painting, statue, poem – is not the work of art. The work takes place when a human being cooperates with the product so that the outcome is an experience that is enjoyed because of its liberating and ordered properties”.

Acknowledgements

This artwork was developed within the “Cultural Adventures” research project, hosted by the University of Technology of Architecture and Design of King Mongkut Thonburi, Thailand, and the Research Center for Arts and Communication (CIAC) at the University of Algarve, Portugal. Conversations with Phanuphong Songkhong played an important role in the early stages of this project. The author would also like to acknowledge the significant contributions from students at the King Mongkut’s University of Technology Thonburi, who played a crucial role in providing the Thai utterances, as well as their translations.

References

- Bonet, Eugeni. 2007. Calculated Cinema. *Verbindingen / Junctions*, 5.
- Dewey, J. (1980). *Art as Experience*. New York: GP Putnam's Sons. p. 213.
- Galanter, Philip (2014). Complexism and the Role of Evolutionary Art. In J. Romero & P. Machado (eds.) *The Art of Artificial Evolution A Handbook on Evolutionary Art and Music*, pp. 311-333. Springer Berlin, Berlin.
- Grba, Dejan (2017). Avoid setup: Insights and implications of generative cinema. *Technoetic Arts* 15(3), pp. 247-260.
- Ishai, A., Fairhall, S. L., & Pepperell, R. (2007). Perception, memory and aesthetics of indeterminate art. *Brain Research Bulletin*, 73(4-6), pp. 319-324. DOI:10.1016/j.brainresbull.2007.04.009
- McDonald, Gail (1993). *Learning to be Modern: Pound, Eliot, and the American University*. Oxford: Clarendon Press. p. 145.
- McLuhan, M. (1988). *Letters of Marshall McLuhan*, eds. M. Molinaro, C. McLuhan, W. Toye. Oxford: Oxford University Press. p. 193.
- Peirce, C. S. (2012). *Collected Papers*. Harvard University Press: Cambridge, MA, USA. pp. 1931-1958.
- Pound, Ezra (1914). Vorticism. *The Fortnightly Review*, 96. pp. 461-471. <https://fortnightlyreview.co.uk/vorticism/>.
- Pound, Ezra (1970). *Guide to Kulchur*. New York: New Directions. p. 51.
- Smith, L. F., Smith, J. K., & Tinio, P. P. L. (2017). Time spent viewing art and reading labels. *Psychology of Aesthetics, Creativity, and the Arts*, 11(1), 77-85. DOI:10.1037/aca0000049
- Singla, N. (2014). Motion detection based on frame difference method. *International Journal of Information & Computation Technology*, 4(15), pp. 1559-1565.
- Springgay, S., Irwin, R. L., & Kind, S. W. (2005). A/r/tography as Living Inquiry Through Art and Text. *Qualitative Inquiry*, 11(6). p. 900.
- Suresh, D. S., & Lavanya, M. P. (2014). Motion Detection and Tracking using Background Subtraction and Consecutive Frames Difference Method. *International Journal of Research Studies in Science, Engineering and Technology*, 1(5), pp. 16-22.
- Veiga, P. A. (2019). A/r/cography: Art, Research and Communication. In *Proceedings of Artech 2019, the 9th International Conference on Digital and Interactive Arts*. DOI: 10.1145/3359852.3359859
- Vrobel, Susie (2011). A Second Eye and Wings for the Cyclops: A Fractal Temporal Analogy to Hockney's Collages. *Acta Systemica*, 11(1). The International Institute for Advanced Studies in Systems Research and Cybernetics, ISSN 1813-4769