Towards an architecture for measuring kid's emotions in a lego education context through data collection and Self Assessment Manikins (SAM) workshops, in Ecuador.

Cabezas Tapia Patrick David[1], Díaz Padilla Danny Sebastián[2], López Rodríguez Juan Antonio[3], Morales Banda Juan José[4], Pantoja Pino Andrés Sebastián[5], Sanango Simbaña Edison Ubaldo[6]

[1−2−3−4−5−6]Faculty of Systems Engineering
Escuela Politécnica Nacional, Quito, Ecuador.
{[1]patrick.cabezas, [2]danny.diaz, [3]juan.lopez08, [4]juan.morales01 ,[5]andres.pantoja, [6]edison.sanango}@epn.edu.ec

Abstract—
Keywords: Kid's emotions, data collection, Microsoft Face Detection API

I. INTRODUCTION

Historically, education has not been a main concern in Ecuador. Because of this, the education system has not been improved nor automatized. Having old methodologies and techniques, this topic had a large gap from international educational levels. Actually, these levels are based on specific software systems and leading technologies. These have not been implemented in this nation.

On the other hand, it is known that every system has to be developed following an architecture. Systems' functionality relies on a well-structured architecture and design on which they are established. It becomes more important when dealing with data such as emotions. Furthermore, when this data come from kids and their experience while learning. Keeping that information in mind, it is key to develop an architecture for data collection. That architecture would be used to measure kids' emotions when learning.

Many architectures involving data collection have been developed in the past. For instance, Perso2U Architecture focuses on emotions in order to drive UI adaptation. Moreover, an architecture that uses facial expressions to recognize emotions has been used before. Nevertheless, adding more information to the architecture with other techniques will help perform data analysis.

The proposed architecture is composed by three general elements: input, processing and output. In the input section gives the data needed. Also, a person can manually record information. This is known as implicit and explicit information respectively. Then, the data source is stored in two locations, on local storage and the cloud. Thus, the data are analyzed by a tool in order to get emotions' values among other attributes. Later, a file is generated with all the information obtained. Finally, the data are sent to the output section where it is presented.

The following section of this work presents the related work regarding architectures similar to the one presented by the authors. Next, the experimental study is shown, where the procedure is explained. Then, the results of the work are displayed. Finally, conclusions and perspectives of the work done are presented.

II. RELATED WORK

A. Data Collection Architecture

The architecture is composed by three general elements: input, processing and output. In the input section, a camera takes a photo every 15 seconds which gives the data needed. In this section, another type of tools like microphones, pulse sensors, force sensors, etc. could be used to get more data but, for own purposes and flexibility, a camera was used. Also, a person can manually record information using observation sheets and SAM workshop. This is known as implicit and explicit information respectively. Then, the photos taken are stored in two locations, on a free hosting service and local storage. The second one is important due to any lost connection with the API or the hosting service.

Thus, the processing section begins, the photos taken are analyzed by a face emotion recognition API (this time with Azure Face REST API) provided by Microsoft in order to get Paul Ekman's emotions values among other attributes like age, gender, hair color, etc. A .json file is generated with all the information obtained. Finally, the data is sent to the output section where it is presented with graphs. This last section is not implemented in this architecture because of the subject content limitation.

B. "The recognition of facial expression using distributed characteristics based on own vectors."

The recognition of facial expression using distributed characteristics based on own vectors and a decision-making

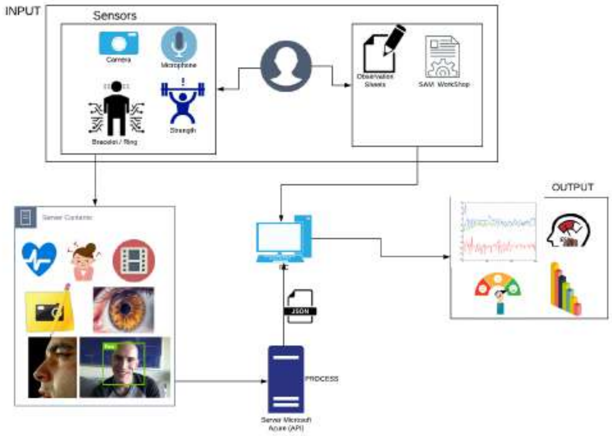

Fig. 1. Global schema of Data Collection Architecture.

technique based on Euclidean distance, by Jeemoni Kalita and India Karen Das, members of the "Department of Electronic and Communications Engineering Assam Don Bosco University Guwahati, India ", is an investigation that proposes a system based on own vectors to recognize facial expressions in digital images.

The proposed problem is based on taking photographs and analyzing them to make a decision that indicates the mood of the person. To solve the problem, the researchers propose the following architecture, showed in Fig. 3. [1]

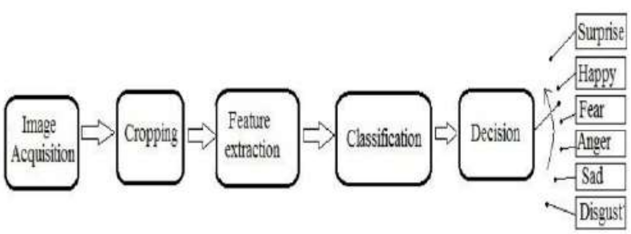

Fig. 2. Architecture for Recognition of Facial Expression Using Eigen-vector

C. *"Considering the aesthetics and usability temporalities in a development process based on models."*

Considering the aesthetics and usability temporalities in a development process based on models, by Sophie Dupuy-Chessa, Yann Laurillau and Eric Céret, is an investigation based on the models of recognition of affection and expression of affection are useful to shape the trajectories of students during focused narrative learning.

This paper also propose to determine the factors affected to the possible pedagogical and motivational benefits of the environments. Students are involved in virtual story worlds, where a three-stage process called an affective loop is recognized.

The proposed architecture is based on a human interacting with a computer, where there are situations with different settings, plot, game activities and empathic pedagogical agents. In these situations we try to predict how the user will act.

The training has been based on tests carried out on a large number of students, where everything is registered in the virtual environment. But this architecture, contrary to the proposal in this paper, fails to fully capture the user's facial expressions or emotions. It is proposed to add cameras and sensors that record the user experience during the interaction with the virtual environment provided by Cristal Island. In the following image, it shows an example of the screen through which the user interacts with the different situations.[2]

Fig. 3. Example of the Cristal Island screen

D. *Perso2U Achitecture*

It is composed by three main components: the Inferring Engine, the Adaptation Engine and the Interactive System. The first one works with the context of Use (users with their emotions, platform, environment), initially, it detects changes in the context of use with external sensors while the emotion wrapper is responsible for providing the emotion values to know if the emotion of the user is positive -this means happiness-, negative -could be anger, contempt, disgust, fear or sadness-, surprise or neutral.

Then, the context of use is sent to the adaptation engine where selects a UI variant and UI parameters to adapt better according to the user. Finally, the interactive system uses the parameters and change the variant so it could be displayed.[5]

E. *"A unifying reference framework for multi-target user interfaces"*

There are different ways to perform the process of recognizing a person's emotions, an example of which is indicated in [3]. In this document describes a framework that serves as a reference for classifying user interfaces that support multiple objectives, or multiple contexts of use in the field of context-sensitive computing.
Where a context is broken down into three end users of the interactive system, the computing platform (hardware and software), and the physical working environment.

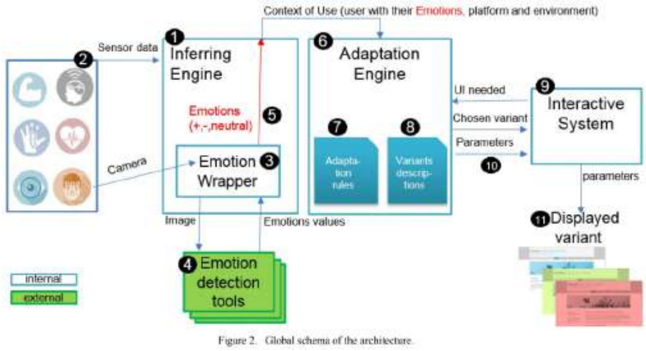

Fig. 4. Global schema of Perso2U architecture.

The difference with the system is that it focuses on analysing people's emotions so that they can later be analysed according to the stage of activity carried out, since the interface is static and does not change due to the user's context, the interface [3], recognises the context and changes according to it.
While the system proposed, tries to generate statistics of what emotions a person experiences when he is in the process of learning or other new activity for the user and how each of the stages can change the emotions of the user.[3].

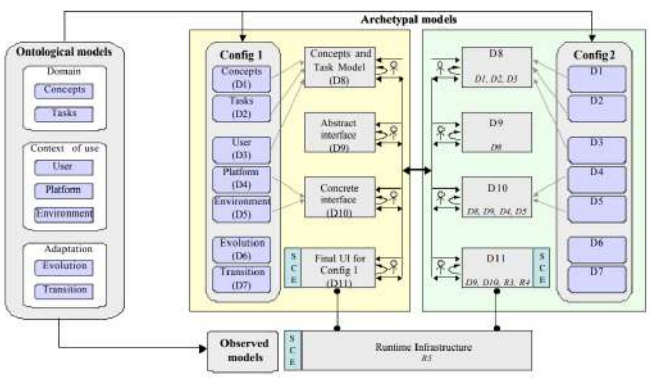

Fig. 5. The Unifying Reference Framework.

F. "Adaptive intelligent user interfaces with emotion recognition"

The objective of creating adaptive user interfaces to facilitate natural communication between the human and the computer through the state of the emotions of the users and to respond to those emotions with a model that adapts to each user are raised in this document, supported by laboratory and virtual reality environments to collect the data necessary for the analysis.

Of discriminant functions that is based on physiological patterns of the users, with a simulator experiment that allows to provoke emotions using the KNN, DFA and MBP algorithms, to recognize emotions with an accuracy between 72 to 84 %, by means of exposure to provoking events (panic/fear, frustration/anger and boredom/sleepiness). Using the formalization of Bayesian Belief Networks to develop the adaptive model.[4]

This architecture uses several algorithms to analyze emotions, but these are used to adapt the user interfaces, but it doesn't make a global analysis of the data to know how the user's emotions changed along a transfer.
Another difference is that Perso2U uses external tools to make the analysis of emotions, with a cloud-based system.

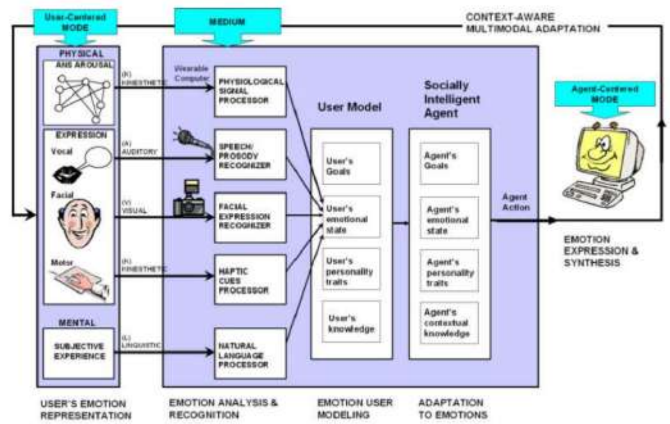

Fig. 6. Human Multi-modal Affect Expression matched with Multimedia Computer Sensing.

Architecture	Face Recognition	Camera	Sensor	GUI
B	YES	YES	NO	NO
C	YES	YES	NO	YES
D	YES	YES	YES	YES
E	YES	YES	NO	YES
F	YES	YES	YES	YES

TABLE I
COMPARISON OF ARCHITECTURES.

G. "Criticism of Related Work"

All architectures perform a Face recognition system to detect emotions, also all use cameras to generate the necessary data for analysis. But the image processing, since in some of the architectures this is done in an external system and others in an internal system, in the case of Data Collection Architecture uses an online Api that receives the images and returns the user is emotions which will later be analyzed.

In the case of sensors, Data Collection Architecture does not use sensors or unlike other architectures, and the adaptive GUI is not contemplated in Data Collection Architecture either, but an analysis of the emotions is implemented according to the stage the user is in.

III. EXPERIMENTAL STUDY

A. Goal and hypothesis

The objective of this experiment is the monitoring of emotions during a collaborative workshop of educational robotics to analyze the learning process based on emotions.To obtain reliable results, it is necessary that the data capture be as accurate as possible. To facilitate the process of data capture cameras are used to record the user experience throughout the experiment, and obtain the user's feelings by applying facial recognition techniques.

If possible, it is also proposed to measure the degree of pupil dilation, which would be an important factor because many studies state that the pupil tends to dilate when people see something they like or feel happy. It is also proposed to use sensors that measure the heart rate, the degree of sweating and the stress levels of the users.

The experiment is divided into three phases, which are to explore, create and share. The objective is summarized in capturing all possible data during the different phases.

So the hypotheses of the experiment are:

H1: It is possible to obtain the different emotions that the user experiences during the experiment performed.

H2: Based on the fact that users learn more when they feel positive emotions, it is possible to monitor the user's level of learning during the educational robotics workshop.

B. Experimental Method

A LEGO WeDo guided project was developed by two groups of students. The project has sorting and recycling as a main topics. Also, each project has tree phases (explore, create and share) with a truck able to sort two objects based on shape as a result. On the other hand, each group of students has four ten years old kids. Every kid provides information for future analysis.

The first type of information registered is explicit information. It is obtained by having children complete a SAM workshop, at the beginning, in the middle and at the end of every phase of the project. The second type of information, implicit information, is acquired by analyzing kids' photos. These photos focus on children's faces for a future face emotion recognition.

Face emotion recognition is performed by Microsoft Azure Face API. Photos are taken and analyzed every 15 seconds, having eight emotions of every face, among other information, as a result. These emotions (anger, contempt, disgust, fear, happiness, sadness and surprise, along with neutral) come as percentages in every face analyzed.

Furthermore, information such about project's phase, group and a photo ID. Also, every photo taken is download locally and uploaded in a server. These data work as a backup for future activities when needed. Finally, a finished LEGO WeDo project and data collected are the results of the method used.

C. Procedure, tasks and participants

Certain prerequisites are needed to start with the experiment, the first one is to create the workshop using one of the Lego WeDo projects, then, the participants must be created and linked to the workshop, as well as a teacher in charge if its is possible, this was done thanks to CRUD (Create, Read, Update and Delete) operations.

The data collection architecture was tested during a collaborative educational robotics LEGO workshop where the main users were children. Twelve persons participated in the experiment, four students responsible for the project as well as the director of it. There were three kids in the first group and four in the other (seven in total). The 10 years-old kids were from 6th grade of 'Verbo' Primary School.

These types of users are usually quite expressive about their emotions. They are very interested in technology, robotics and workshops, such as the one proposed for this experiment, during these types of activities they usually have fun, which is important to defend the hypothesis that while people have more fun, they also learn more.

Participants showed great interest during the three phases of the experiment (explore, create and share). During the exploration phase, participants were asked a problem. They had to explore the materials provided and find a solution to the problem. The second phase started when participants began to build and implement the solution they defined during the previous phase. Once the construction of the solution was completed, it is time for the exchange phase. In this last phase, the participants expose the built and functional solution, and explain how this helps to solve the problem posed in the first phase.

As said before, in each phase of the project, every kid had to complete a SAM workshop prior to photo capture. During all the three phases, the group of kids were in front of the camera and if possible offering a direct view of themselves. For both of the groups the photos were taken every 15 seconds.

Their emotions were detected with the Microsoft Azure Face API using the photos, also, in case of network problems, those photos were stored locally in each computer for it posterior use with the API.

D. Interaction Data and Images Correction

User pictures were sent to Microsoft Face Detection API (emotions detection tool) for getting the recognized emotions values.

In terms of interaction data, participants' head shots were taken every 15 seconds in real-time. One user's observation represents a data tuple: timestamp, seven emotions (happiness, contempt, anger, disgust, fear, sadness, surprise, neutral), age and gender.

It is important to emphasize that in a photo, 0 or more users may appear, in the case that there are 0 users, a post analysis is performed to clean the images that do not have a face; in the other case for each user a tuple is obtained, that is, for a photo there can be several tuples and this is stored in JSON files.

E. Real-time data visualization

At any moment during the workshop, data visualization play a main topic in educational context. Even though the aim of this experiment is to obtain data during the workshop, it is important to note that some strategies could be made starting from data visualization results in order to improve the learning and teaching process. Thus, the user can take actions if necessary during the execution of the workshop. Moreover, the user can see how the child or children change their mood during the workshop.

The graphs that can be generated are (happiness, contempt, anger, disgust, fear, sadness, surprise, neutral), all these data are extracted from the processing of the Microsoft Face Detection API, these data are graphed and exposed according to the need of the user. Additionally the user can start and stop the workshop at any time, which has a stopwatch to indicate how long the workshop takes.

F. Display of stored data

When the user needs to analyze the data that was recorded at the time the workshop was running. The children on whom the analysis will be performed can be selected as well as their emotions (happiness, contempt, anger, disgust, fear, sadness, surprise, neutral), and the stage he is in (create, explore, share).Data are stored in a web service to be analyzed when necessary, to facilitate the task of research.

G. Measures and method

Paul Ekman introduced a universal emotions set. Those emotions were implemented on many face detection software. Hence, Microsoft Face Detection API considers Paul Ekman's emotions as well as gender, age, color hair, accessories among other attributes and features.

This tool was used since it counts with a free trial option. Even though it has some limitations which could have improved the data analysis, the tool was enough for this experiment. Due to the aim of this experiment it is important to keep in mind that emotions are the principal and the most valuable data. Obviously, there are data like age or face landmarks whose values will help during the data cleaning process. Therefore, Fig. 7 summarize some attributes obtained from Microsoft API.

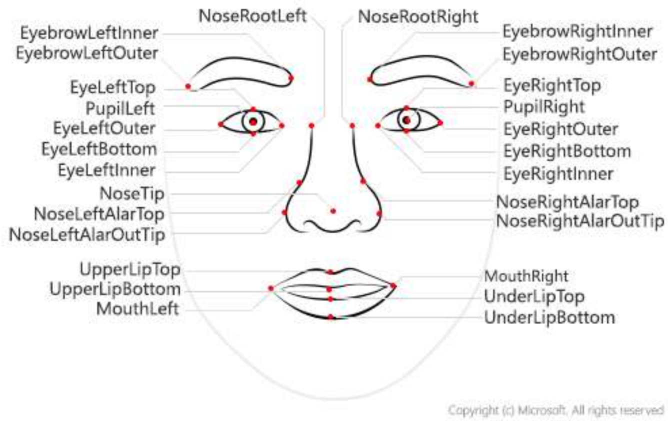

Fig. 7. Face Landmarks returned by Microsoft API.

In order to capture reliable data, it was required to comply some requirements imposed by Microsoft API. Consequently, the images taken must be send in a established format in this case was .png format. Also, the image size must not be higher than 4 Mb. Also, the photo dimensions must be between 36 x 36 to 4096 x 4096 pixels. On the contrary, any photo won´t be able to obtain its data. Provided that those requirements had been satisfied, the input data are able to be processed by the API.

Furthermore, it was necessary to add some attributes in order to have a better post-classification. In addition to Microsoft API attributes, a 'kidId' was added in order to recognize who owns the data. Also, a 'phase' and a 'phaseName' were added, this was useful to split the analysis of the workshop in specific phases. Otherwise, it would not be possible to recognize the end and the beginning of each phase and the photos which belongs to a specific phase.

As stated before, there were two groups in this experiment which causes to add an extra attribute called 'group'. This attribute is manually established depending on the group in which the photo was taken. This value just can be 1 or 2.

At the same time, the 'photoId' was added this is important because all the photos taken follow a sequence which is established with the 'seasonId' whose value is the date and time expressed in milliseconds. Also, the phase name and the number of the photo in that phase. For example, "sesionid_1578920298096Explorar_foto_1.png". In the same way, the url, where the image was saved, was added. It means the hosting url.

All those data were exported to a '.json' file which is recommended if necessary to import to a database such as MongoDb. The Fig. 8 shows an example of a register of a face which was recognized by Microsoft API.


```
{
  "faceId": "7b6ca55e-d4cd-4fa4-bff7-430eee360f18",
  "top": 265,
  "left": 75,
  "width": 96,
  "height": 96,
  "smile": 0.003,
  "pitch": -3.1,
  "roll": -5,
  "yaw": 56,
  "gender": "male",
  "age": 12,
  "moustache": 0,
  "beard": 0,
  "sideburns": 0,
  "glasses": "NoGlasses",
  "anger": 0,
  "contempt": 0,
  "disgust": 0,
  "fear": 0,
  "happiness": 0.003,
  "neutral": 0.676,
  "sadness": 0.001,
  "surprise": 0.32,
  "blurLevel": "medium",
  "blurlevelValue": 0.47,
  "exposureLevel": "overExposure",
  "exposureLevelValue": 1,
  "noiseLevel": "low",
  "noiseLevelValue": 0.05,
  "eyeMakeup": false,
  "lipMakeup": false,
  "accessories": [],
  "foreheadOccluded": false,
  "eyeOccluded": false,
  "mouthOccluded": false,
  "bald": 0.07,
  "invisible": false,
  "hairColor": [
    {
      "color": "brown",
      "confidence": 0.98
    },
    {
      "color": "black",
      "confidence": 0.96
    },
    {
      "color": "other",
      "confidence": 0.24
    },
    {
      "color": "red",
      "confidence": 0.23
    },
    {
      "color": "gray",
      "confidence": 0.11
    },
    {
      "color": "blond",
      "confidence": 0.08
    }
  ],
  "phase": 1,
  "phaseName": "Explorar",
  "group": 1,
  "kidId": 0,
  "photoID": "sesionid_1578920298096Explorar_foto_1.png",
  "imageURL": "
  https://deusgallet.000webhostapp.com/recursos/AAL_emotions/Grupo1/Explorar/sesioni
  d_1578920298096Explorar_foto_1.png"
},{
```


Fig. 8. Example of data obtained from a photo through Microsoft API.

IV. RESULTS

A. Feasibility of an architecture to measure UX for learning (learn+do) Lego education

Considering the time assigned for each phase (ten, sixty and ten minutes respectively) the data collection architecture generated the next results for the first group of children:

- 33 photographs in the exploration phase.
- 203 photographs in the creation phase.
- 54 photographs in the sharing phase.

In other hand, these are the results for the second group of children:

- 29 photographs in the exploration phase.
- 230 photographs in the creation phase.
- 36 photographs in the sharing phase.

It can be seen that the first group has less amount of photos in the creation phase and it is because they finished before group 2, a similar case can be noticed for the sharing phase where the group 1 used more time than group 2.

All of these photos were used in the Microsoft API and stored in a single json file per group, the first one has almost 30 thousand lines and the second around 37 thousand lines with the information explained in previous sections. The variation between these two files is due to how many times the API recognized faces and generated the data.

With this in mind, the data collection architecture is viable to measure user experience in a collaborative educational robotics Lego workshop.

B. Failed emotion recognition

Although Microsoft Face Detection API can recognize different attributes from a photo, it is necessary the photo complies all the requirements said before. Otherwise, it will not be able to obtain the required data for this experiment. In addition, there are some factors which affect the results. For instance, when a kid put a hand in their jaw or mouth it is quiet probable that Microsoft API could not give any face recognized. Therefore, no data was available. On the other hand, the experiment took photos every 15 seconds. This time could change, it could be less but, because of API limitations it was established in that range.

However, emotions were not the only attribute which obtained wrong values from Microsoft API. In some cases, the gender attribute could not identify its value with a good accuracy. For instance, the same boy was recognized as a man in some pictures but in others as a woman. A valid explanation for this event, is the age of the kids. All the participants had 10 years old. At that age, not all their face features are well-formed.

There are varied reasons for having a failed emotion recognition. It also depends of the hardware used, camera's resolution, even internet speed for the calling to Microsoft API. Therefore, it is crucial to consider all the variables which can affect the results. Also, the following experiment limitations describe some points that have to be reviewed in order to success in this experiment.

C. Experiment limitations

Despite shown results, there are some limitations that need to be taken into consideration. In the first place, only two methods for emotion recognition are used. Many other methods can be used for a better emotion recognition. Second, Microsoft API usage depend on the conditions of the photo taken.

The place where the experiment took part, light, illumination and focus can affect the photo taken. Next, the experiment used children that played in groups while obtaining data. Having this information in mind, social environment can affect SAM workshops results and kids' movement can reduce photographs' quality. Finally, the methodology implemented relies heavily on hardware and network. It is known that these elements can fail during experimentation and data analysis.

V. CONCLUSIONS AND PERSPECTIVES

REFERENCES

[1] J. Rowe, B. Mott, S. MCQUIGGAN, J. Sabourin, S. LEE, and J. LESTER, 'Crystal Island: A Narrative-Centered Learning Environment for Eighth Grade Microbiology', 14th Int. Conf. AI Educ. Workshop Proc., Jan. 2009.

[2] S. Dupuy-Chessa, Y. Laurillau, and E. Céret, Considering aesthetics and usability temporalities in a model based development process. 2016.

[3] G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt, 'A Unifying Reference Framework for Multi-Target User Interfaces', Interact. Comput., vol. 15, pp. 289–308, Jun. 2003, doi: 10.1016/S0953-5438(03)00010-9.

[4] F. Nasoz, 'ADAPTIVE INTELLIGENT USER INTERFACES WITH EMOTION RECOGNITION', University of Central Florida Orlando, Florida, 2004.

[5] E. Céret, S. Dupuy-Chessa, and J. Galindo, 'Toward a UI Adaptation Approach Driven by User Emotions', Mar. 2017.