

**Alma Mater Studiorum – Università di Bologna**

**DOTTORATO DI RICERCA IN**

**Ingegneria Elettronica, delle Telecomunicazioni e  
Tecnologie dell'Informazione**

**Ciclo XXVIII**

**Settore Concorsuale di afferenza:** 09/E3 - ELETTRONICA

**Settore Scientifico disciplinare:** ING-INF/01 - ELETTRONICA

**ADVANCED TECHNOLOGIES  
FOR HUMAN-COMPUTER INTERFACES IN MIXED REALITY**

**Presentata da:** Marco Marchesi

**Coordinatore Dottorato**

**Relatore**

Prof. Alessandro Vanelli Coralli

Chiar.mo Prof. Bruno Riccò

**Esame finale Anno 2016**

*To my parents*



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>The Mixed Reality paradigm</b>	<b>7</b>
<b>3</b>	<b>BCI Learning: BRAVO</b>	<b>13</b>
3.1	Brain-Computer Interfaces for interactive and immersive experiences . . . . .	14
3.1.1	BCI-based Learning Systems . . . . .	15
3.2	BRAVO . . . . .	15
3.2.1	Design Architecture and Sw/Hw Implementation . . . . .	16
3.2.2	A Cultural Heritage Demo . . . . .	19
3.2.3	Single Mode . . . . .	20
3.3	Evaluation . . . . .	22
3.3.1	A first Qualitative Test . . . . .	22
3.3.2	Discussion . . . . .	26
3.4	Computer-based assessment . . . . .	27
3.4.1	Computerized Adaptive Testing . . . . .	28
3.5	A possible Collaborative Mode Approach . . . . .	32
3.6	Future Work . . . . .	33
<b>4</b>	<b>Capturing reality: Augmented Graphics</b>	<b>35</b>
4.1	The object recognition problem . . . . .	35
4.2	Moment Invariants theory . . . . .	36
4.3	Object Detection . . . . .	39
4.3.1	Pre-process image . . . . .	40
4.3.2	Find contours . . . . .	43
4.3.3	Process Contours . . . . .	44
4.3.4	Keypoints Extraction . . . . .	46
4.3.5	Shape Matching . . . . .	48
4.3.6	Create Mask . . . . .	48
4.4	Implementation . . . . .	50



4.5	Evaluation . . . . .	51
4.5.1	Discussion . . . . .	53
4.6	Future Work . . . . .	53
<b>5</b>	<b>Hand Motion in Virtual Reality: GLOVR</b>	<b>55</b>
5.1	Related Work . . . . .	56
5.2	Overview . . . . .	57
5.2.1	Design Architecture . . . . .	59
5.2.2	Implementation . . . . .	67
5.2.3	Evaluation . . . . .	71
5.2.4	Quantitative Method . . . . .	73
5.2.5	Qualitative Method . . . . .	76
5.2.6	Discussion . . . . .	76
5.2.7	Future Work . . . . .	78
<b>6</b>	<b>Conclusions</b>	<b>81</b>
6.1	Discussion . . . . .	81
6.2	Future Work . . . . .	83
6.3	Acknowledgements . . . . .	83
	<b>Bibliography</b>	<b>85</b>

Keywords:

*Human-Computer Interfaces*  
*Mixed Reality*  
*Wearable devices*

# Chapter 1

## Introduction

We are surrounded by data. Information that can be encoded, compressed and manipulated in many ways. As human beings, we trust our five senses, that allow us to experience the world and communicate. Since our birth, the amount of data that every day we can acquire is impressive and such a richness reflects the complexity of humankind in arts, technology, etc. How did this mechanism evolved through centuries? We can assert for sure that as long as progress goes ahead, the ability of humans to get useful information from the surrounding world has increased exponentially. Fundamental discoveries arisen from the observation of nature speed up this process. But observation is not only a task that involve human senses, it has been enormously enhanced with the help of *artificial* senses. Think about how the telescope allowed Galileo to give birth to observational astronomy and microscope was fundamental for Louis Pasteur for putting foundations to microbiology field. In the 20th century Quantum Physics revolutionized the concept of “measure” while the advent of computers and the consequent progress in AI showed how large amounts of data can contain some sort of “intelligence” themselves. Machines learn and generate a superimposed *layer* of reality.

How data generated by humans and machines are related today? To give an answer we will present three projects where we considered data fundamental in creating solid connections between what we intend as “Reality” and what has emerged definitively in the last two years as its extension, the “Virtuality”. Such context of *Mixed Reality* will be our playground.

This document is organized as follows:

*Chapter 2: The Mixed Reality Paradigm* will give an introduction to the *Virtuality Continuum*, the concept in which Virtual and Augmented Reality lie. We will see how the Reality and the Virtuality are two extremes more and

more connected as long as data increase. The next chapters are dedicated to three applicative examples of such an ideal space where Reality and Virtuality can co-exist: *Chapter 3: A BCI Application: BRAVO* will present *BRAVO*, an e-learning tool based on the user's brain activity recorded by a single channel EEG headset, its architecture and the evaluation results; *Chapter 4: Capturing reality: Augmented Graphics* will give an overview of the object detection task and how is currently managed. As a fast and cheap alternative to sophisticated training-based algorithms, we will introduce *Augmented Graphics*, a framework particularly suitable for mobile applications. We will describe how it works and the evaluation results with two datasets of images; *Chapter 5: Motion Sensing in Virtual Reality: GLOVR*, will describe a wearable hand controller designed for Virtual Reality environments. It uses inertial sensors to offer directional controls and recognize gestures and it features a microphone for implementing a natural language service, a solution that expands the opportunities to interact with external voice-assisted applications. In *Chapter 6: Conclusions* we will summarize some of the most significant results obtained by our research and will guess the future trends in Human-Computer Interfaces and Mixed Reality environments.

## Chapter 2

# The Mixed Reality paradigm

The definition of *Mixed Reality* was given for the first time in 1994, as “*anywhere between the extrema of the virtuality continuum*” [Milgram and Kishino, 1994], where *Virtuality Continuum* refers to earlier definitions of *Mediated Reality*, coined by Steve Mann in his pioneering research in wearable computing [Mann and Fung, 2001].

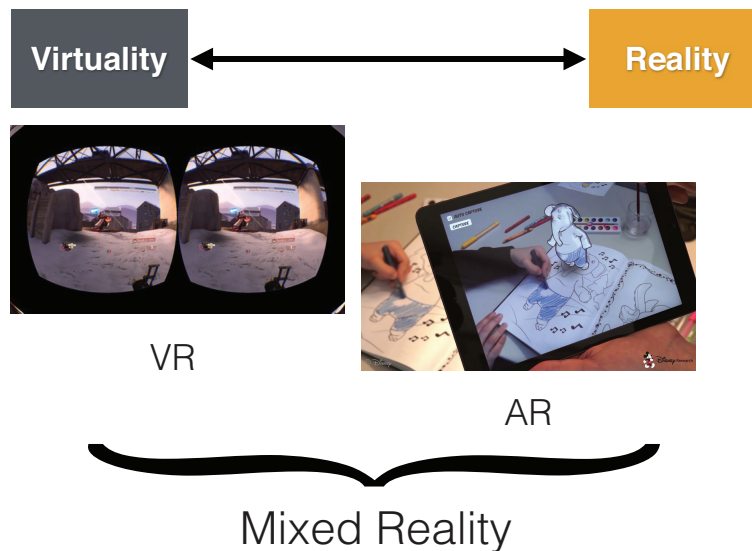


Figure 2.1: The Virtuality Continuum line that goes from Virtuality to Reality and back. Virtual Reality (VR) and Augmented Reality (AR) are the main technologies involved in this revolution.

Where Virtuality and Reality are different? In terms of human senses, so far the most evident distinction lied in visual experience. In Reality what



Figure 2.2: The evolution of Steve Mann's work through two decades of pioneering in wearable computing.

we see is what we get from the surrounding environment by means of our visual system, while *Virtuality* is what commonly is considered as digital *computer-generated* graphics where depth perception and perspective rules are simulated in a flat space given by one or more displays. Audio sources can be generated by computers, and they occupy the *Hearing* sense as long as they dominate on the surrounding ambient sound. Also, *Touch* simulation have been proposed in several research projects with the challenge of deliver a intuitive sense of tactile feedback like with CyberGrasp [G. Nikolakis, 2004] or other recent solutions [Cameron et al., 2011]. Thus, if we restrict the human senses to just the common five ones (actually it's a matter of debate in scientific community), only three of five, better to say two and a half over five<sup>1</sup>, are simulated.

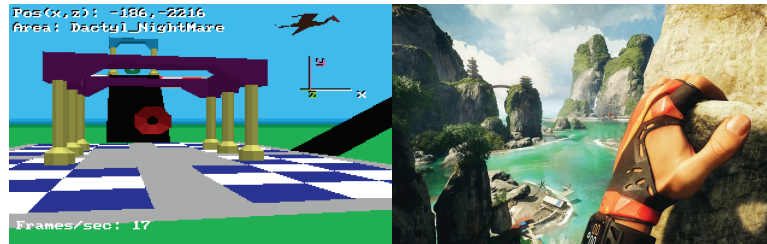


Figure 2.3: A comparison of VR graphical level of rendering in the 1993 and in 2015.

As long as the quality of visual and audio output reach higher levels year after year, in terms both of resolution and realism, the feedback that

<sup>1</sup>Consider that *Artificial Touch* does not offer yet a realistic user experience.

can be acquired from the user is continuously increasing as well. Since the introduction of sensors in mass market, user data acquisition has involved dramatically an elevated amount of devices, faster, cheaper, smaller, more accurate, able to record “knowledge” from the real world and the users and convert in digital information.

How to manage such a huge amount of bits? Parallelism, Heterogeneous Computing, Cloud Computing, are just a few concepts that have arisen in the last decade and represent the ground level architecture that gave significance to Machine Learning and all the data-related fields. Since any *data science* is based on the information that is found on nature or is computer-generated, and subsequently any user experience in digital world is influenced by some sort of outputs given by processed data, the fundamental question that goes through the chapters of this document is:

*What is the role of data in the Mixed Reality continuum?*

In first experiments of VR, users faced some issues that appeared insurmountable at that time: low resolution environments, headsets with a limited field of view, a significant latency. In latest generation of VR, devices are ubiquitous, users can ideally play scenarios in any place: at home, on metro, at school, on an airplane, giving origin of unexpected user experiences that are enriched by the data acquired from biosignals, gestures, tracking.

In AR, the combination of real and virtual images has challenged researchers and engineers back since the 1960’s [Rekimoto and Nagao, 1995] [Thomas, 2012] and [Billinghurst et al., 2015].



Figure 2.4: The “Sword of Damocles”, the first Head-Mounted Display by I. Sutherland and D.Sproull [Sutherland, 1968].

Thus, did the concept of reality change through decades of technologic progress? We don't think it did. But, without going to any philosophical discussion, we can say for sure that the amount of information that is shared between Reality and Virtuality in the Virtuality Continuum line has changed drastically.

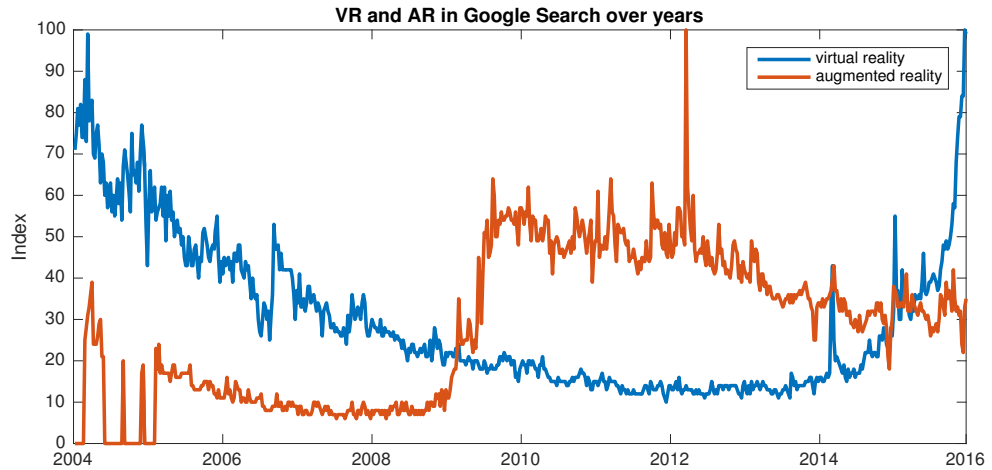


Figure 2.5: Trends in “virtual reality” and “augmented reality” terms in Google Trends [Google, 2016]. The index ranges from 0 to 100 where “100” stands for the maximum normalized value of popularity in Google searches.

The information that is acquired from the Reality by means of sensors and then is encoded digitally in data is partially transmitted in Virtuality. Conversely, what the users can do in Virtuality can influence their real experience, creating a *Virtuality Continuum Feedback*. Let us do some examples:

- In a VR game, player’s performance can be significantly influenced by motion sickness issues due to the graphics latency. The psychophysiological effects of navigation in VR can be recorded as changes in EEG, ECG and EMG activity [Kim et al., 2001] and used as a correction factor for reducing cybersickness;
- children can transform their colorful hand drawn creations into digitally animated models and play with them [Magenat et al., 2015];
- real-time control of three-dimensional avatars is an important topic in the context of computer games and significantly gets information from the human movements [Lee et al., 2002]

In the following, we will discuss three examples of “mutual influence” between real and virtual environment and how Virtuality is “feed” by the amount of data acquired by sensors that provide a digital perspective of the real world, while human senses and then Reality receive continuous feedback from the virtual world. The three applications have been designed for different systems, from mobile devices to desktop, according to the goals that we have planned to reach, but ideally are portable on any modern mobile/desktop platform.

- *BRAVO* [Marchesi and Riccò, 2013b] is a e-learning mobile application that records EEG brain activity in order to change the learning path according to user’s attention and stress levels. As learning task can involve any sort of contents, we developed a user interface that integrates textual contents, images and 3d models, in order to guarantee an high engagement. After a first design approach, we performed some evaluation tests with public audience and then we have a proposed a new solution, based on Computerized Adaptive Testing; BRAVO was not designed as a VR/AR application in strict sense, but its exploration of 3d models can be easily ported to AR. Furthermore, brain activity is exploited passively, user doesn’t have “to think about what he/she is thinking”, and thus it reinforces the idea of virtual worlds and digital contents influenced by human senses;
- *Augmented Graphics* (AG) [Marchesi and Riccò, 2013a] is an object recognition framework based on moment invariants theory. Differently from the training-based recognition methods, AG runs an elaborated shape matching process to detect a sample object in the real environment. Once done, the recognized object can be exported and superimpose digital graphics in external applications, realizing the so called *Augmented Virtuality* concept;
- *GLOVR* [Marchesi and Riccò, 2016] is a wearable hand controller that integrates inertial sensors and a microphone to extend user interaction outside the VR environment and combine them in a robust multimodal input. While playing with a VR scene, the user can send voice messages with natural language and a gesture recognition system let the avatar do actions.





## Chapter 3

# BCI Learning: BRAVO

Education is undergoing a large transformation due to the fast evolution of digital technologies, providing new tools for both teachers and students. *Learning Managing Systems* have been introduced in primary schools as well as in university courses, while the dramatic increase of internet connections allows *Massive Open Online Courses* (MOOC) to grow, some of which are supported by interactive and social networking technologies [Coursera, 2016]. In this scenario, Brain-Computer Interfaces can play a significant role [Nijholt et al., 2008] in education, as they can provide useful information about student's attention and motivation [Rebolledo-Mendez et al., 2010] and enhance the learning curve according to the emotional information classified. On the other hand, mobile devices, such as tablets and smartphones, are widely recognized as suitable complements of conventional learning tools, because of their widespread utilization and their extremely powerful interfaces. In this context BRAVO (Brain Virtual Operator) has been designed as a system for content visualization in a mobile e-learning application. BRAVO makes use of the brain activity acquired by the BCI (particularly, attention and meditation levels). This permits to know which parts of the content are most difficult for the user, so as to propose them in the most appropriate form, in a different way or with a reduced or deeper level of difficulty. Although primarily dedicated to educational purposes, the system can be easily adapted to other applications, such as any interactive experiences where a lack of attention expressed by the user can be a significant issue to be solved.

### 3.1 Brain-Computer Interfaces for interactive and immersive experiences

In the last two decades Brain-Computer Interfaces (BCI) became a popular research topic thanks to the progress in computers and electronic equipments and the increased understanding of brain functionality [Wolpaw et al., 2002]. BCI research showed novelty on finding alternative ways for people with disabilities to communicate and physically interact, but since the first attempts it was clear the need to create a set of controlled conditions that ensure safe and reliable tests in various contexts. Various ideas of what an “immersive” environment can improve the brain learning have been proposed. The work of Lecuyer and others [Lecuyer et al., 2008] suggested some paths in research on BCI and VR, proposing BCI as a substitute of common hand controllers and gamepads, but also as an input that change the VR content, according to the user’s brain activity. Other works demonstrated how three-dimensional virtual environments guarantee better user’s response than 2D ones [Leeb et al., 2007].

A field where BCI found wide application in terms of “immersivity” has been that one of *serious games* [Liarokapis et al., 2014]. Researchers aim to exploit EEG activity to “fully control an avatar” and “examine the reaction of users while playing the game” [Liarokapis et al., 2013], and many of them using simple dry electrode headsets that are suitable for testing with a large number of players [Yoon et al., 2013]. Single EEG channel devices are partially responsible for the rising popularity of BCI in general audience [Crowley et al., 2010] but they show some significant limits in terms of the accuracy with which cognitive processes like attention can be measured. However, they have been subject of several BCI-based mobile applications as well [Coulton et al., 2011]. We will see how BCI on mobile will be the context where we decided to design BRAVO.

Furthermore, how the right source of attention can be recognized and selected in an environment full of “noise”? In these terms any efficient immersive experience aims to generate an environment where the user’s engagement can lie. The *gaze point* is where the user is looking at. In a VR environment, under the assumption that the gaze point is directed to the attention source, various visual attention models have been proposed, by means of eye tracking systems [Courgeon et al., 2014] or avoiding them [Lee et al., 2009] [Hillaire et al., 2012].

### 3.1.1 BCI-based Learning Systems

Brain-Computer Interfaces appeared as a promising candidate in measuring attention and cognitive effort, thanks to the considerable progress in capturing EEG signals with less or non-intrusive solutions. Attention is the cognitive process that most clearly can be an indicator of the learning process and its efficacy [Rebolledo-Mendez et al., 2010]. Most of the classification methods for recognizing cognitive patterns are based on Support Vector Machines (SVM)<sup>1</sup> that give better results [Liu et al., 2013] than  $k$ NN classification [Li et al., 2011]. Other methods like Hidden Markov Model (HMM) were used to infer engagement in students [Beal et al., 2007]. Not only EEG, also functional magnetic Resonance Imaging (fMRI) equipments have been exploited [Anderson et al., 2011].

Research on adaptive systems have suggested that the estimate of attention (or engagement) and workload from EEG performs well in giving prediction on learning progress. Galan and Beal used a 9 sensor EEG headset to acquire signals then processed to produce classification of mental states, such as Engagement, Distraction, Drowsiness and Cognitive Workload, that allowed to predict problem outcomes sensibly better than chance [Cirett Galán and Beal, 2012].

## 3.2 BRAVO

BRAVO starts from the results given by two previous projects: *Mobie*, a graphical tool for creating interactive videos where the story was influenced by the user's brain activity [Marchesi et al., 2011], and *NEU*, an virtual environment editor, where each element (characters, illumination, game mechanics) took in account the trend of the user's attention for changing over time [Marchesi, 2012]. Both researches were conducted by means of a one-channel EEG headset, that compared with other sophisticated multichannel devices, showed its advantages in terms of usability, thanks to the placement of pre-amplified dry electrodes instead of gel-based ones typical of lab setup.

---

<sup>1</sup>Briefly, SVMs aim to find a hyperplane that separate the data points of different classes with the maximum distance.



Figure 3.1: Mindwave: the single-channel EEG headset under test.

BRAVO has been design primarily for e-learning purposes. For this reason at first version was developed as a client application of *Moodle*, a open source tool for creating courses, lectures and assignments through customizable modules [Moodle, 2016] [Jin, 2012]. A Moodle course works as follows:

- A teacher adds resources and activities for their students, varying from a single page to a complex set of tasks where learning progress is tracked in a number of ways and indicators: progress bars, checklists, engagement analytics report, individual learning plans, course status trackers;
- The students enrolled in a course can have grades, submit assignments and can also be added to groups if tasks to assign need to be differentiated.

Moodle architecture appeared limited in terms of real-time capabilities and mobile implementation. BRAVO extends Moodle features and allows working with any kind of contents, though particularly suitable for those situations that need interactive and navigable presentation, because of their higher level of engagement. A critical point for the effectiveness of the system is the correct estimate of the students interest and motivation. To this purpose, touchable interfaces and mobile devices can help in better tracking the user activity. Consequently, BRAVO features many elements of *gamification*, such as progress bars, flags and scores as they appear immediately familiar to the audience [Deterding et al., 2011]. In practice, with BRAVO the learning process can be monitored in real-time by means of the BCI, in order to customize the contents to the student's learning curve.

### 3.2.1 Design Architecture and Sw/Hw Implementation

BRAVO has been conceived as a client/server application for mobile devices. Data are requested dynamically according to the assessment and the current

brain activity, acquired by the BCI.

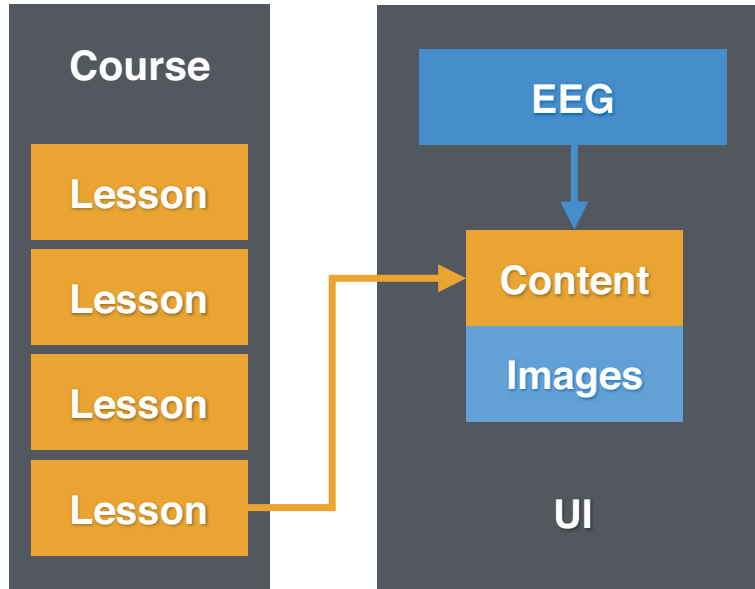


Figure 3.2: BRAVO server/client. Contents are the result of a data selection based on the brain activity.

As one of the primary goals was to obtain a portable learning tool, at first glance some specifications were considered as important features:

- a mobile device connected to the web for downloading the media contents from a server;
- a wireless EEG headset connected to the mobile device via Bluetooth;
- a server that stores all the information taken from the BCI and the learning process.

For testing BRAVO, Neurosky’s *Mindwave* headset has been chosen, because of its ease of use with mobile systems. Despite its simplicity as BCI, Neurosky technology has been implemented in several consumer applications and research projects [Folcher et al., 2014] as well as previous experiments in measuring students’ attention while reading [Mostow et al., 2011].

BRAVO acquires high-level brain patterns every second, in the form of *Attention* and *Meditation* levels, in a range 0-100, by means of a Neurosky’s algorithm called *eSense*, implemented in the ThinkGear chip. After a very short setup, the device starts detecting EEG 0.5-100Hz brain spectrum with

a frequency up to 512Hz, from which it calculates attention and meditation levels every second [Rebolledo-Mendez et al., 2009]. Basically, that values are in relationship with alpha and beta and gamma waves.

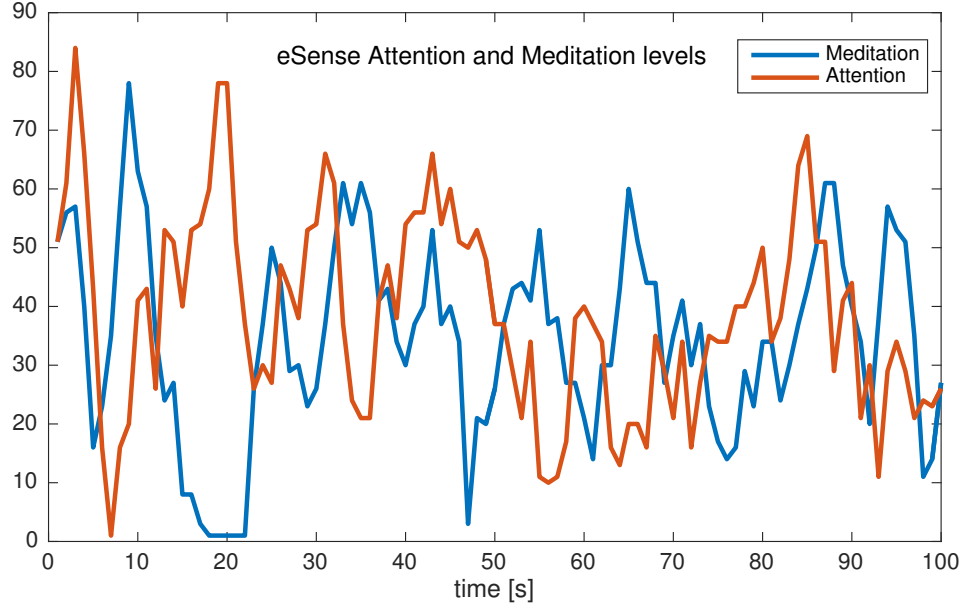


Figure 3.3: A recorded session of Attention and Meditation levels.

At this stage, we preferred to work with the eSense algorithm, instead of analyze EEG data, because we wanted to focus more on the design of the tool and how the user should interact properly with his/her own EEG signals. The attention and meditation levels acquired look enough accurate for our study and give enough information on how the user evolves. How the learning contents can vary depending of the user's brain activity is not trivial. For example, a detected low attention level may suggest the presence of difficulties with previous parts of the learning program, or a temporary lack of interest in it. Thus, such parts may be repeated or an easier approach to the topic may be proposed in order to go ahead in the study session and get attention. On the contrary, well-focused students are stimulated to work harder and reach higher learning results at the end. In BRAVO contents are showed with variable complexity according to an *Ability Level*(AL), that increases/decreases once the last sequence of brain levels has been processed.

### 3.2.2 A Cultural Heritage Demo

First demo was based on an extension of the capabilities of Moodle, in order to offer interactive courses for single users or groups of students. The topic of the course was “Historical Monuments in Bologna” and we choose seven famous places rich of History and Art to represent our city. To generate more engagement, together with simple text and images, we implemented also a 3d view that showed the monument as a 3d textured model, kindly offered by CINECA.

Each monument offers a series of *Hot Spots*, interactive points of interest that show the levels of ability reached by the user for each content. Once a user touches the Hot Spot, the content related to the point of interested is loaded on the screen, according to the Ability Level estimated by BRAVO in it (see Fig. 3.4).

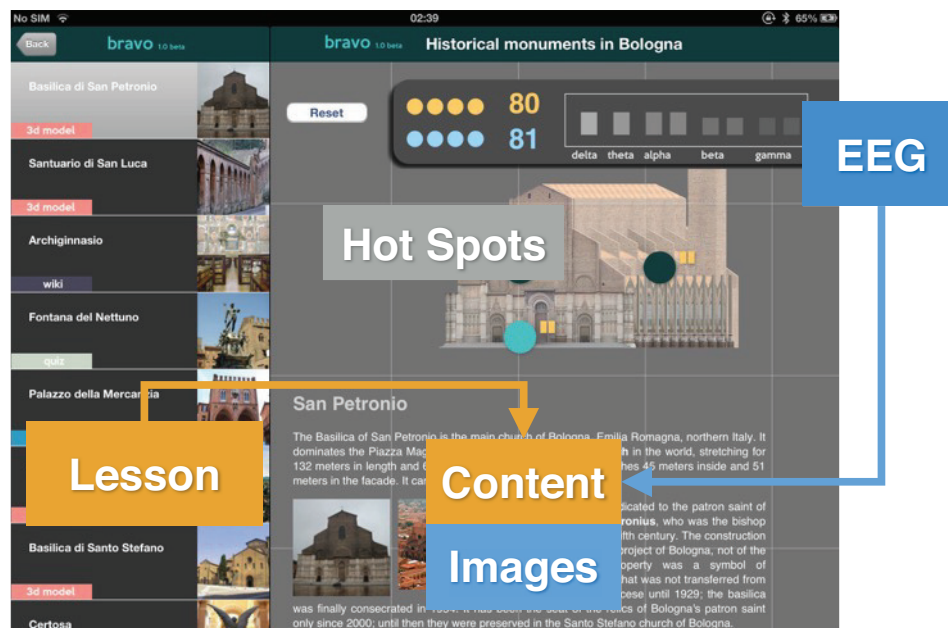


Figure 3.4: BRAVO GUI. *On the left:* the selection menu. *On the right:* the content view, on top the BCI Bar shows the EEG spectrum and the Attention and Meditation levels; centered, the content is a 3D model used to navigate through the Hotspots, each of them show the levels the user has reached so far.



### 3.2.3 Single Mode

In *Single Mode* users study and do their assignments alone, without the collaboration of others. Such a continuous feedback has been initially exploited with the introduction of a *Threshold Approach*(TA).

#### Threshold Approach

Both A and M levels have been divided in four intervals, according to Neurosky's guidelines, and the median calculated in the last  $n$  seconds is taken in account. Median value has been preferred to mean because is more robust to the presence of outliers that occur in brain signals.

80-100	80-100
60-80	60-80
40-60	40-60
0-40	0-40

Figure 3.5: Categories for eSense attention and meditation levels.

TA assumes that the probability of acquiring a certain ability reflects directly the brain activity measured by the headset and is directly dependent on the values recorded in attention and meditation:

$$P(\theta_i) = f_i(A, M) \quad (3.1)$$

where  $f_i$  depends on the specific item. For instance, examinees who are good in math calculus need low attention and show high levels of meditation while they try to solve math problems, whereas from low ability levels usually we expect more mental effort and stress (that is, lower meditation levels). In other tasks the attention needed to pass the assignment or, in any case, "record" properly the content, can be different. Thus, TA works as following (Fig. 3.6):

1. A lesson made of textual and media contents is loaded;
2. while the user is going through the lesson, A and M are recorded;

3. once the user stops or if he/she interacts with other elements in the interface, the task is stopped;
4. TA calculates median values of  $A$  and  $M$  and compares them to the threshold, updating the complexity of the argument according to the  $AL$  reached by the user in the lesson under study. If  $A(M) > 40$  the user is promoted to the upper level, if  $A(M) > 60$  he/she promoted to the upper level with *bonus* = 1 and if  $A(M) > 80$  with a *bonus* = 2. Conversely, for  $A(M) < 40$  and  $A(M) < 20$  the bonus is decreased by 1 or 2. The “bonus” is an additional term to set the new  $AL$  that takes into account how far the median is from average;
5. bonus and  $AL$  are updated and then the content according on them.

Consider the following example:  $bonus = 2, AL = 3$ . The assignment is stopped and  $A < 40$ . In this case it is expected that the user’s Ability Level has to be decremented by 1, but the bonus can balance the decrement of  $AL$  up to its maximum value. In this case after the bonus application:  $bonus = 0, AL = 3$ .

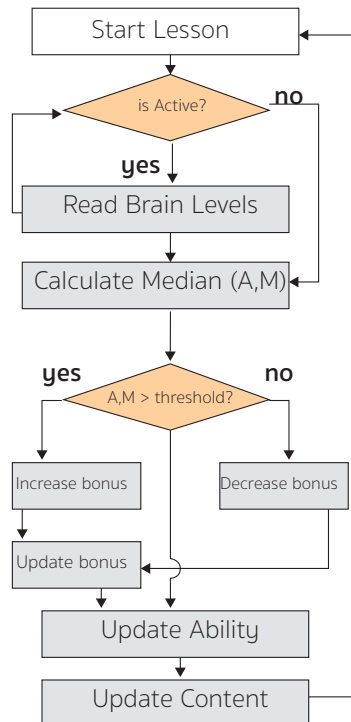


Figure 3.6: Diagram of flux of BRAVO threshold algorithm.

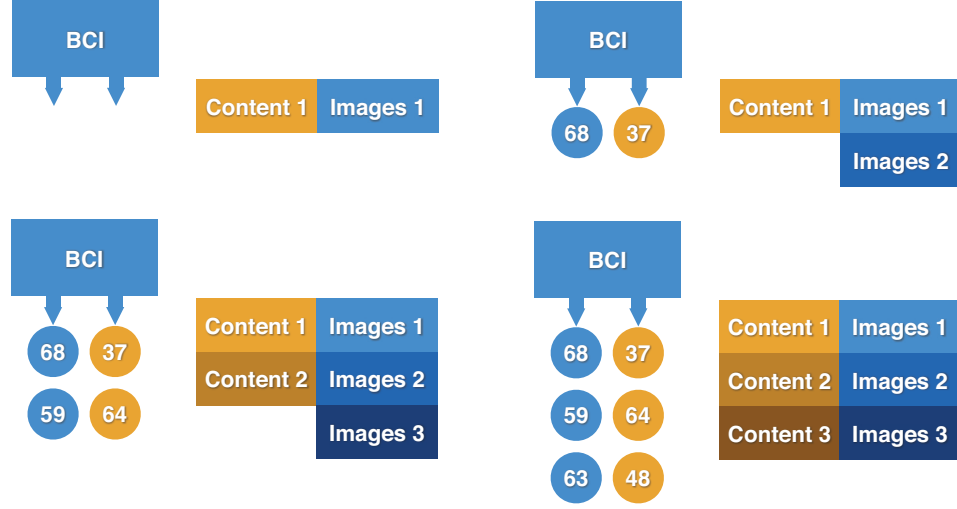


Figure 3.7: An example of single mode with median level approach. *Top Left*: the lesson starts with basic textual content and images. *Top Right*: high meditation level adds visual contents while low attention level keeps textual contents at the same complexity. *Bottom Left*: a deeper focus on the content is followed by an addition in textual content, as well as the insertion of new images. *Bottom Right*: an average level of attention causes an increment of content. In this case, because of a progress bonus.

### 3.3 Evaluation

As we assume that the learning process is enhanced by means of a neuro-feedback, tests on lab and with public audience has been executed.

#### 3.3.1 A first Qualitative Test

A first test was made in a Cultural Heritage event (ArcheoVirtual 2012) with the help of 28 participants (15 male, 13 female) from different countries in Europe, with an age in the range 16-48 (mean 29.29). Such range, even if it is wide, it groups people with an age that is considered as producing a similar “adult” EEG activity and, consequently, similar cognitive processes [Kellaway, 1990]. All the participants didn’t have any previous experience with BRAVO. Asking their occupation was important to identify any possible relationship between the level of knowledge reached as function of the brain activity.

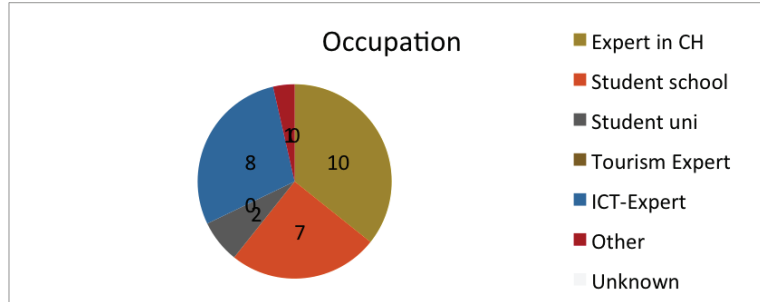


Figure 3.8: The occupation of BRAVO testers:

From Fig. 3.8 approximately one third of the testers consider him/herself an expert in Cultural Heritage (CH), thus we can guess should be relevant in some way. Large majority of the testers were very confident with tablets (46.43%) and interactive technologies. For the evaluation, they had to wear the EEG headset and navigate one of the CH contents proposed in a iPad tablet for a time comprised between 5 and 10 minutes, an interval of time that we considered adequate for a public event. Users chose one of the seven monuments from the list, selected a hot spot on the 3D model and then started reading from the simplest content to to the higher, according to their Ability Level they reached. During the recording session, their brain levels real-time graphs were hidden, in order to limit the sources of attention in the tablet screen.

After the session, a balanced Likert scale test on a 7 point agreement has been proposed with the following 12 opinions about the overall experience with BRAVO (See Table 3.1): ArcheoVirtual organizers elaborated three other scales, in a range from -3 to 3: *Perspicuity* ( $mean = 1.54$ ), *Efficiency* ( $mean = 1.36$ ) and *Stimulation* ( $mean = 1.05$ ), showing the positive impact of BRAVO as a new possible interactive learning experience.

Five additional multiple choice questions was asked, whereas *functions* we intended:

1. Navigation on list of monuments
2. Selection of Hot Spots of the monument
3. Fruition of text and images
4. Visualization on demand of the brain levels in real-time

For Q1 the 96.45% of the users found all functions “useful and interesting” while just 3.57% (one person) judged some functions “unnecessary”. For Q2 the 80.77% considered all functions available and for Q3 the 57.14% thought

Survey Statements	
not understandable	understandable
organized	cluttered
motivating	demotivating
inefficient	efficient
complicated	easy
boring	exciting
clear	confusing
fast	slow
not interesting	interesting
easy to learn	difficult to learn
valuable	inferior
impractical	practical

Table 3.1: The survey proposed to all the participants after the session with BRAVO. The Likert scale goes from 1 to 7.

Survey Statements	
Q1	What do you think about the functions of BRAVO?
Q2	Was every function available that you wish to have?
Q3	What do you think about the handling of BRAVO?
Q4	What do you think about the time it took to use it?
Q5	What do you think about your experience with BRAVO?

Table 3.2: General impressions about BRAVO asked to the ArcheoVirtual testers.

about BRAVO handling as “easy to use from the beginning” whereas the 35.71% “first hard, then easy”. Time spent (Q4) was considered “adequate” for the 89.29%. The whole experience (Q5) was rated “totally new” for the 78.57%.

We analyzed the recorded brain levels for all the 28 participants together with the qualitative feedback. Three of the participants didn’t provide relevant answers and have been excluded.

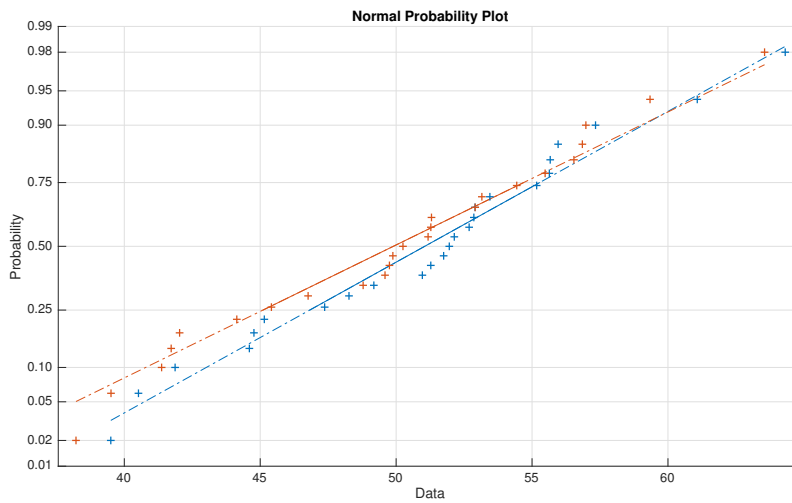


Figure 3.9: The median values recorded are normally distributed.

Within the questions proposed, we selected a subset of four Likert scales, corrected to the same increasing scale 0–7: *boring/exciting, clear/confusing, easy to learn/difficult to learn, not interesting/interesting*. For each tester, the sum of the four scales represents what we called an *Experience Score* (ES). Assuming ES as an index of the learning experience, our hypothesis was: the higher the ES the higher the brain levels recorded during the “learning” session. From each participant we calculated the median and the correlation coefficients between ES and the attention (A) and meditation (M) levels. From Fig. 3.10 we see ES can be considered in relationship with A and M (that are not correlated each other) for most of the participants.

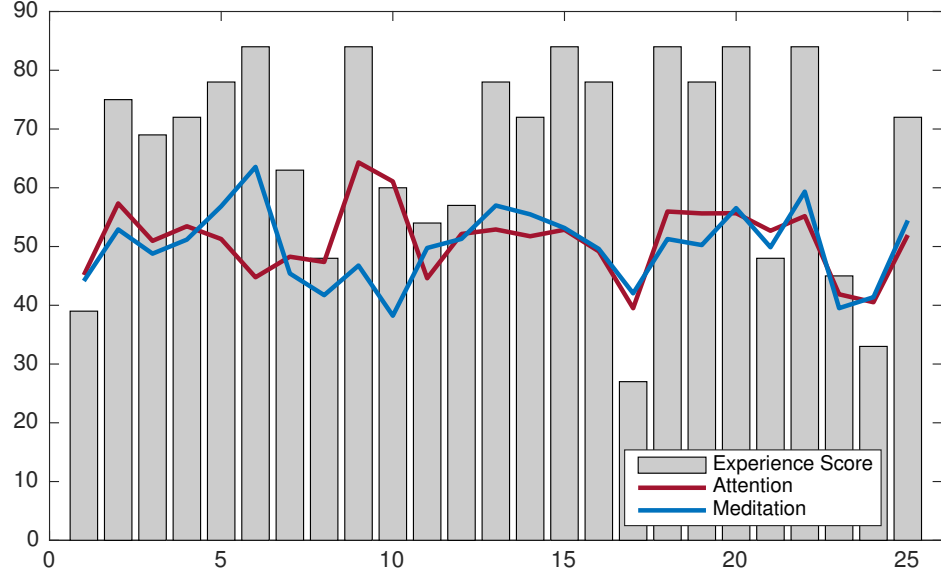


Figure 3.10: The *Experience Score* calculated among 25 participants in the range 4-28.

From the correlation coefficient matrix for A:

$$\begin{array}{cc} CC1 & = 1.0000 & 0.6804 \\ & 0.6804 & 1.0000 \end{array} \quad (3.2)$$

while for M:

$$\begin{array}{cc} CC2 & = 1.0000 & 0.7345 \\ & 0.7345 & 1.0000 \end{array} \quad (3.3)$$

and the  $p\text{-value} < 0.05$  for both, we identify them as significant correlations between the interest expressed in the experiment and the attention recorded from the brain activity.

### 3.3.2 Discussion

From the tests taken with public audience we realized how interactive systems can potentially enhance the learning process. A significant correlation between the brain signals and the user experience show how the A and M levels can give further information about how much the user will learn proficiently, starting from the assumption that “high focus and low stress” is a pre-condition of a productive learning session. However, testing in a public

environment pointed out one of the intrinsic limits of Brain-Computer Interfaces: how the different “targets of attention” can be distinguished each other? In a public environment people that talk, noises and sounds at distance, displays, moving objects, all can contribute to the overall attention in time. With this concern in mind we strongly considered to take advantage of “gamification” elements to help the user to keep focused.

We designed Threshold Approach with the goal to give a solution that improve the overall user experience by means of the interpretation of two cognitive processes, attention and meditation, under the assumption that such advantage could be a benefit for the user’s learning curve, that fits a customizable content selection.

In the following section we will see how to implement the information acquired from the brain activity to a traditional *Computerized Adaptive Testing* methodology.

### 3.4 Computer-based assessment

In the following section will be introduced Adaptive Testing as one of the key elements that characterized the further design of BRAVO. The first individual tests were introduced by Alfred Binet (1857-1911), a French psychologist and educator that classified the test items according to their level of difficulty. Its adaptive form took in account also the personal data such as the age of the examinee. The advent of computers was fundamental for the development of more complex psychometric models, such as the Two-Stage Adaptive Testing or the Stratified Adaptive Test [Weiss, 1985]. Around 1950’s, in contrast to the Classical Test Theory, it appeared the Item Response Theory (IRT, also know as Latent Trait Theory), thanks to the work of Frederic M. Lord [Lord, 1980]. In a IRT-based test each item is not equally difficult. To do so, IRT assumes that all items are locally independent and the response of an examinee can be modeled by a *Item Response Function*:

$$P(\theta_i) = c_i + \frac{1 - c_i}{1 + \exp(-a_i(\theta_i - b_i))} \quad (3.4)$$

Where  $\theta_i$  are the abilities of the examinee that are to be measured.  $a_i$  is the discrimination vector. It measures the variation from the low probability to the high probability region. With a good  $a_i$  the probability that a low ability examinee answers correctly is low where the probability for a high ability one is high.  $b_i$  is the item difficulty, whereas  $c_i$  represents the “guess”, that is the probability that a low ability examinee gives a correct answer.



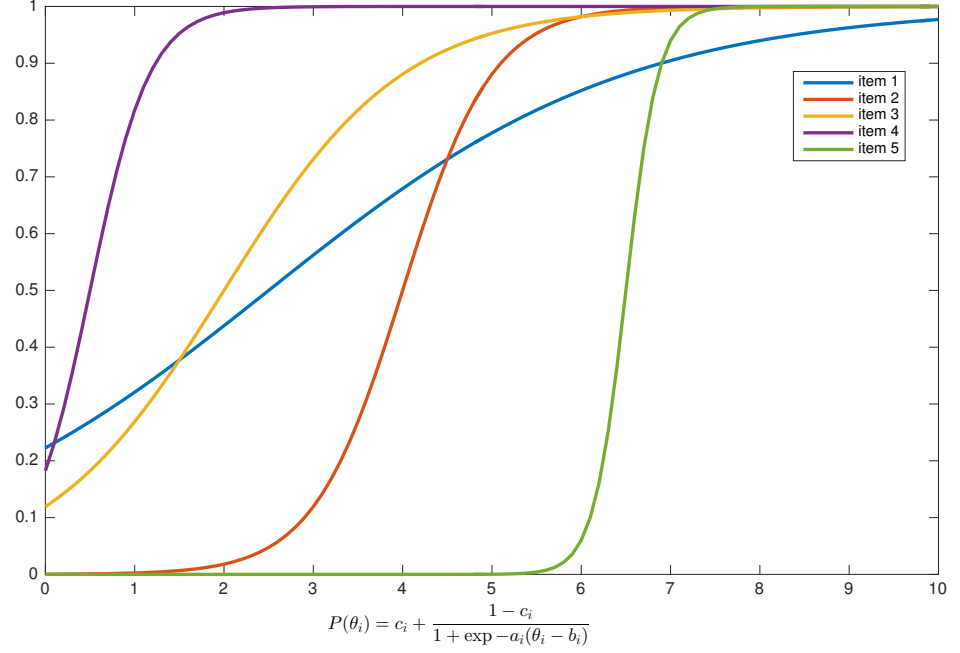


Figure 3.11: The sigmoid curve for a few items.  $c_i = 0$  (guess).

### 3.4.1 Computerized Adaptive Testing

Among the implementations of IRT, the Computerized Adaptive Testing (CAT) methodology is the most popular [Reckase, 1974]. CAT is an iterative algorithm that aims to maximize the precision of the exam according to the current estimate of the examinee's ability. Typically a CAT test performs better for testers with high or low levels of abilities and gives high precision with a lower amount of questions. CAT is based on an *Item Bank*, that is, a set of items (larger than the maximum number of items for a test) that customize the test depending of the tester responses. From that, a starting point has to be choose in two different ways: i) considering some prior information about the examinee or ii) without any prior information and assuming an average ability and therefore items of medium difficulty [Parshall et al., 2012].

The goal of each CAT iteration is to measure the examinee's ability according to the current item. This is achieved by Maximum Likelihood or Bayesian methods. Particularly the latter ones, that assume a prior knowledge of the ability, are considered more robust [Segall, 1996]:

$$f(\theta|u) = L(u|\theta) \frac{f(\theta)}{f(u)} \quad (3.5)$$

where  $u$  is the vector of the examinee's answers,  $f(u)$  is the marginal probability of  $u$  and  $f(\theta)$  is the prior distribution of the ability and  $L$  is the *likelihood function*. The item that maximizes the information with theta is selected [WEISS and KINGSBURY, 1984]. The concept of "information" can be Local or Global [Piton-Gonçalves and Aluísio, 2012] and its definition is related to the concept of the *precision* with which a parameter is estimated. Following the *Fisher's Information* definition [Cam and Yang, 2000] we can introduce the *Item Information Function* as<sup>2</sup>:

$$IIF = \frac{1}{\text{variance}_{\theta}} \quad (3.6)$$

CAT algorithm runs until a stopping criterion is satisfied because one of the following targets has been reached:

- i. the available items in the item bank
- ii. A determined number of items
- iii. The target accuracy needed to prove the examinee ability
- iv. A lower threshold in ability

In CAT every item is selected with the goal to maximize the ability  $\theta_i$  How the brain activity can jointly affects the parameters of an item response?

### A novel design: CAT based logistic Approach

Johns and Wolf proposed an Item Response Theory model with the application of HMM to infer student's motivation, based on three behavioral states: *motivated*, *unmotivated-guess* and *unmotivated-hint* [Johns and Woolf, 2006]. We have seen how in Threshold Approach we assumed that the probability to reach a determined ability can be predicted in some way from the brain activity levels. A second approach that we followed with BRAVO was to consider attention and meditation as parameters of a logistic function, in analogy of what happens with CAT methodologies:

$$P(\theta_i) = \frac{1}{1 + \exp(-a_i(\theta_i - b_i))} \quad (3.7)$$

---

<sup>2</sup>For a dichotomous response in a 1PL  $I(\theta) = P(\theta)(1 - P(\theta))$ .

That with the explicit contribution of A and M levels gives:

$$P(\theta|B) = P(B|\theta) \frac{P(\theta)}{P(B)}, B = f(A, M) \quad (3.8)$$

where  $P(B|\theta)$  is the probability of a brain activity  $B$  given  $\theta$ , according to the Bayesian's Rule. According to this, BRAVO acts as a “predictor”: it aims to predicts the results (that is, the item difficulty level) that a CAT algorithm should give, adding the contribution of the user's brain activity, that it is assumed to be representative of the capability to answers correctly to an assignment (see Fig. 3.13).

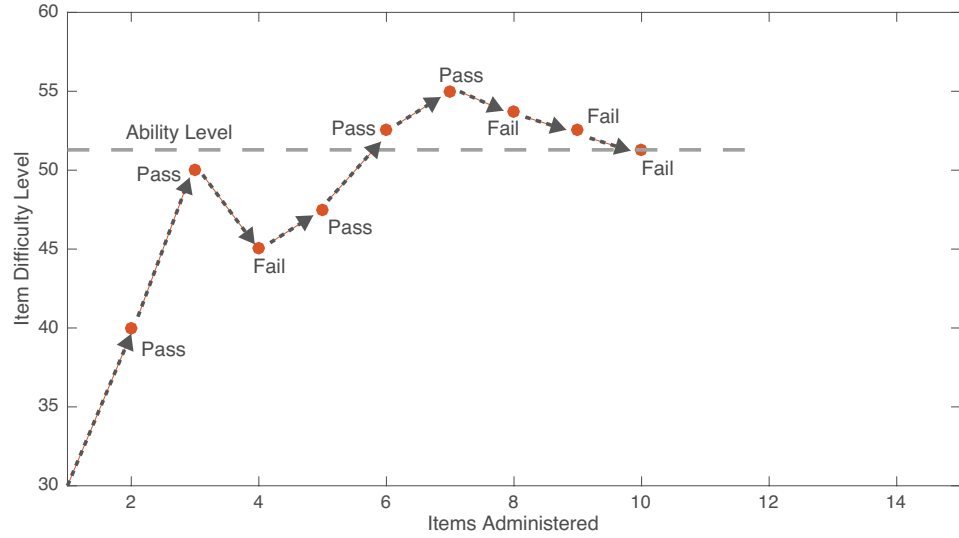


Figure 3.12: A dichotomous CAT test administration. The algorithm tends to converge to an ability level with less items than a non-adaptive test.

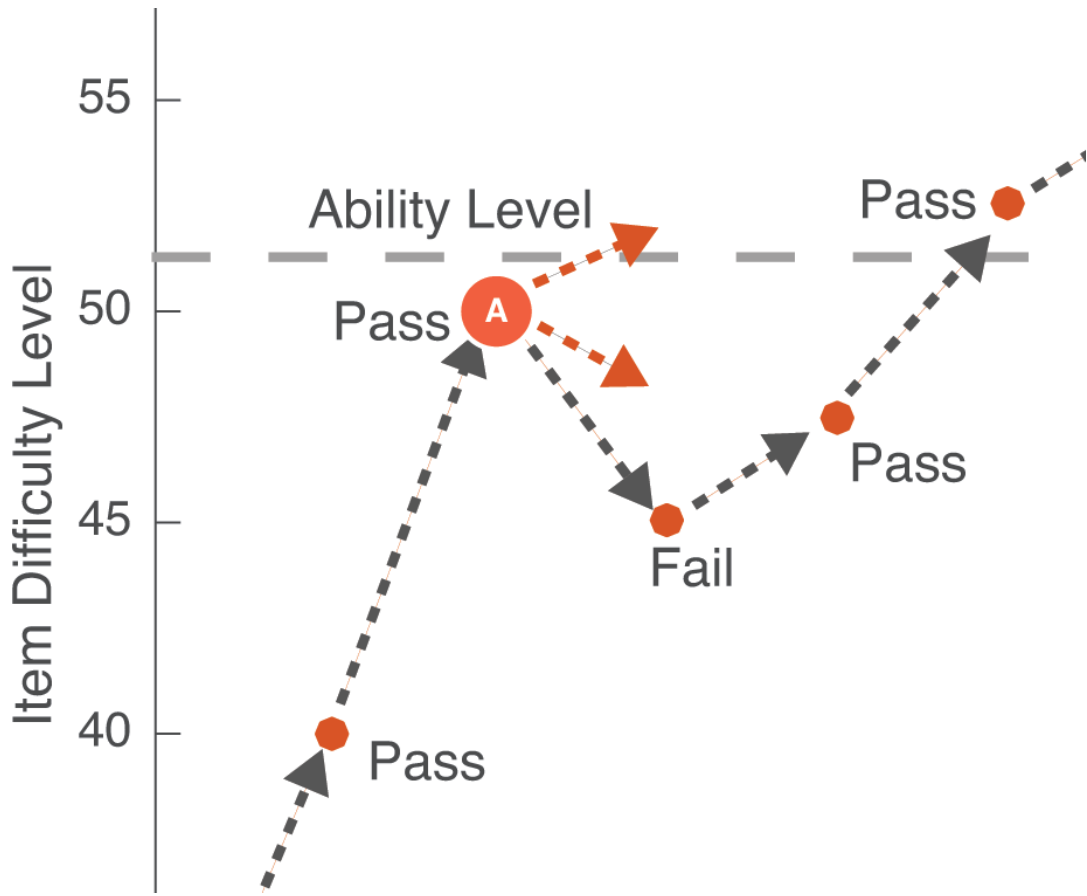


Figure 3.13: CAT corrected by BRAVO by means of the attention level can yield to better predictions than that ones achieved by the “traditional” version.

A basic CAT can benefit of the additional information given by the brain activity in terms of attention and meditation levels. The following tentative *Neuro Computerized Adaptive Testing* (NCAT) is proposed:

---

**Algorithm 1** A NCAT algorithm

---

- 1: **while** Reading brain activity **do**
  - 2:   **repeat**
  - 3:     Select item from the bank as function of the ability  $\theta$ ,  $I(\theta)$
  - 4:     Show item to the examinee
  - 5:     Update ability  $\theta$  as function of the last response
  - 6:     Update ability  $\theta$  as function of the brain activity,  $\theta = f(A, M)$
  - 7:   **until** Criterion termination is met
  - 8: **end while**
-

Basically we expect that low meditation with a determined  $\theta$  will decrease the probability to answer correctly to  $I(\theta)$  whereas a high meditation or high attention can be symptoms of a correct response. In fact, our hypothesis is that a significantly high brain levels recorded in the act to answer to a item can potentially increase the probability of answer correctly also the future item. Thus, our guess is that a slightly higher level of ability  $\theta$  can be suggested for the next question.

Summarizing, our design proposal suggests four alternatives, given  $\theta_I$  and  $\theta_{II}$  the increments in ability due respectively to the correct answer and the brain activity hypothesis.

Brain Activity Prediction	Answer	Update
V	V	$\theta \leftarrow \theta + \theta_I + \theta_{II}$
V	F	$\theta \leftarrow \theta + \theta_I - \theta_{II}$
F	V	$\theta \leftarrow \theta - \theta_I + \theta_{II}$
F	F	$\theta \leftarrow \theta - \theta_I - \theta_{II}$

Table 3.3: The possible alternatives in our Neuro-CAT design.

### 3.5 A possible Collaborative Mode Approach

In *Single Mode* the item difficulty was partially inferred from the brain activity of a single user. Despite the fact the online assessment is individual, it is interesting to understand how assignments can be solved collaboratively and how to model such a group modality. Let us assume a group of four students. Each student's ability score is based on the responses to an assignment. Every student receives the same items<sup>3</sup> with the fundamental difference that has to solve them collaboratively, thus the final group score will be based on the sum of the just one response for each item. The students have two practical ways to decide the responses:

- i. each student solves a fraction of the test separately
- ii. for every item all the students compare their tentative results and decide together which is the most probable response

For both cases we need to assume that all the students previously acquired the same level of ability / knowledge. On the contrary, students with low

---

<sup>3</sup>What we expect from a Classical Test Theory [Novick, 1966].

levels of ability could affect negatively the total score. As we consider the acquired ability a function of the brain activity, as in (3.7), thus the study group can guess from monitoring the brain activity if one of the students will have some difficulty to give a valid response. In this case, a “quick” solution may be to weigh each student’s contribution (ii) or redistribute items already assigned to the students that show a higher ability (i).

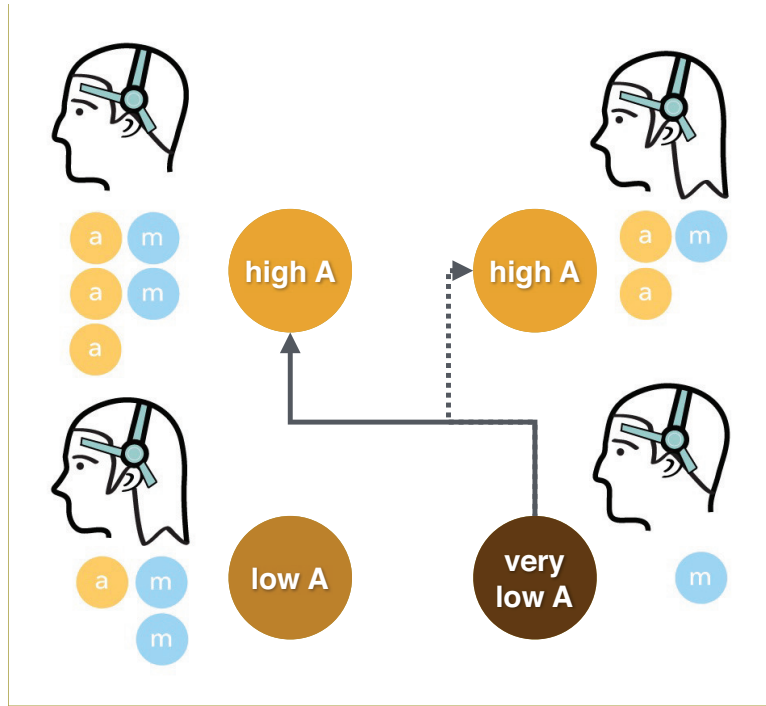


Figure 3.14: BRAVO Collaborative Mode: the student at the bottom right shows low levels of attention and meditation. The two students at the top can help him to solve his part of assignment.

## 3.6 Future Work

We introduced a new learning system that exploited brain signals to predict user’s learning performance and customize the content to offer, with the goal to maximize the learning task efficiency. Starting from previous works in BCIs applied to education, and in Computerized Adaptive Testing, we designed two different approaches that add content generality to the prior research, since it is applicable to various forms of contents and learning methodologies. Then, we tested a first approach, demonstrating how engagement can be recognized in the brain levels acquired by a simple EEG headset

in use.

Several applications aim to give more data, exploiting inputs other than brain signals to classify user's attention, as merging eye tracking data [Ghiani et al., 2015]. However, we think that the next generation of EEG headsets with dry electrodes will allow to run reliable tests outside the "lab setup", offering a wider range of cognitive patterns to be classified on mobile devices via cloud services [Castellani et al., 2014]. In this context, we will be able to design concrete solutions for optimizing the learning process in terms of the user's brain performance.

## Chapter 4

# Capturing reality: Augmented Graphics

Conventionally, in Augmented Reality (AR) a real environment is enriched by means of virtual graphical elements. On the contrary, in Augmented Virtuality (AV) a virtual scene is mixed with objects or people taken from the reality [Hughes et al., 2005]. The potential superposition of real actors to a mixed environment has been extensively explored [Charles, 2004] and interfaces between virtuality and reality have been investigated [Koleva et al., 2000]. Moreover, Computer Vision algorithms have been applied to MR in gaming [Hammond, 2008], where images acquired by a camera were analyzed to recognize objects able to interact with the game elements. Smartphones are ideal for enjoying a fictional narration augmented by real objects 1) taken by the embedded camera, 2) analyzed by image processing algorithms and then eventually 3) placed in the virtual scene. All the steps are executed in the same portable device in all kind of environments.

### 4.1 The object recognition problem

Object recognition is a popular topic in Computer Vision and research in that field has tested several solutions for multiclass recognition and face detection problems. Particularly the first one, it is strongly dependent from the huge amount of data needed to train the system: ImageNet is made by more than 14 million images organized according to the WordNet hierarchy [Deng et al., 2009].

In the field, a wide variety of scenarios can be solved focusing only on the shape of the object to find, based on the presence of a number of keypoints and their spatial configuration. In this case the most critical task to be



solved is the correspondence between the sample and the test keypoints. Correspondence is usually solved by means of descriptors, like in SIFT [Lowe, 2004], in Shape Contexts [Belongie et al., 2002] or in Geometrical Blur [Berg and Malik, 2001]. In geometric hashing [Lamdan et al., 1990], a vote model is implemented, but no concrete correspondence is given. Decision trees have been used for discriminating between different spatial configurations [German et al., 1997]. Shape classification has been performed using the inner-distance [Ling and Jacobs, 2007], edge-based features [Mikolajczyk et al., 2003], Low Distortion correspondences [Berg et al., 2005] or Content-based image retrieval algorithms [Lillo et al., 2010]. With Curvature Scale Space the contours are sectioned in convex and concave curvatures by means of a multi-scale analysis [Mokhtarian et al., 1996].

Shape recognition is generally categorized into contour-based and region-based descriptors [Bober, 2001]. The first are extracted by the object boundaries and eventually divided into segments, called primitives. Region-based descriptors, conversely, take in account the internal information of an object. In this way, they can describe complex objects discriminating by the internal data. Sketch recognition can be considered a similar problem, it has been studied extensively in the past and several solutions have been proposed, most of them focused on specific domains. *Ladder* [Hammond and Davis, 2007] was a sketch description language where the user had to write a sketch grammar for each new domain. From the origin version by 2003 other versions such as *PaleoSketch* [Paulson and Hammond, 2008] were released. A Query-adaptive Shape Topic model was proposed by Sun [Sun et al., 2012] for applications such as sketch tagging, image tagging and sketch-based image search.

## 4.2 Moment Invariants theory

The goal of image recognition is to generate robust image patterns, based on descriptors that should have invariance properties, despite the fact that images are typically corrupted by noise, deformations and occlusions. Local descriptors [Mikolajczyk and Schmid, 2005] have proven successfully in applications such as object recognition [Ferrari et al., 2006], data mining [Sivic and Zisserman, 2003], face and texture analysis [Hadid et al., 2014], image categorization [Mele et al., 2006]. Among them, moment invariants are extensively used for feature extraction in a wide range of imaging applications.

Moments in Computer Vision are related to the pixel intensity, as follows:

$$m_{ji} = \sum_{x,y} (array(x,y) \cdot x^j \cdot y^i) \quad (4.1)$$

A particular combination of moments is the *Mass Center*, given by  $\bar{x} = \frac{m_{10}}{m_{00}}$ ,  $\bar{y} = \frac{m_{01}}{m_{00}}$ , that represents a *centroid* of the geometry (see Fig. 4.1).

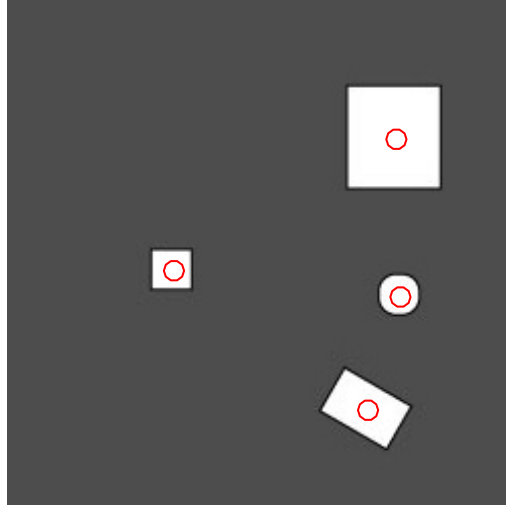


Figure 4.1: Centroids (red circles) calculated for basic shapes refer to the “spatial” center of mass.

For our context we are interested in Normalized Central Moments, in the form:

$$\eta_{ji} = \frac{\mu_{ji}}{m_{00}^{\frac{i+j}{2}+1}} \quad (4.2)$$

The moments and the related invariants have been originated from the theory of algebraic invariants by David Hilbert in the 19th century [Hilbert, 1994], but was only Hu [Hu, 1962] in 1962 that introduced the famous seven moment

invariants:

$$\begin{aligned}
I_1 &= \eta_{20} + \eta_{02} \\
I_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
I_4 &= (\eta_{30}\eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
I_5 &= (\eta_{30} - 3\eta_{12})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
&\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})^2 \\
I_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
&\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - \\
&\quad 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\
&\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{4.3}$$

Moment invariants have been proved to be invariants to the image scale, rotation with the assumption of images considered with infinite resolution and noise-free. Furthermore, projective moment invariants exist in a form of infinite series of moments with positive and negative indices [Suk and Flusser, 2004]. In case of digital images the moment invariants may vary with image geometric transformation and introduce error. Quantitative analysis of such a error have been provided and it has been showed how the fluctuation in moment invariants can be decreased with image spatial resolution [Huang and Leng, 2010]. Hu's Moments have been compared with other invariants, like Fourier Descriptors [Chen et al., 2004] that take in account the pixels along the image contours and need less spatial resolution. It has been proved that moment-base recognition systems performances are compromised in case of object symmetry. For example, all odd-order moments of a center-symmetric object are equal to zero. A new set of invariants robust to symmetry was introduced by Flusser and Suk [Flusser and Suk, 2006]. We will see how with AG we found an alternative *multi-level* solution for the symmetry problem. The idea of using moment invariants as a way to measure similarities between shapes have found in the past decades several implementations, [Dudani et al., 1977] [Yuan and Hui, 2008], using Artificial Neural Networks [Wahi et al., 2012] or SVM for classification of object and non-object data [Nigam et al., 2013]. The shape or the contour of an object is a good basis for invariant recognition, then the massive exploitation of contours detection for that purpose [Sluzek, 1994] [Wei and Wu, 2013]. Furthermore, the identification of a signature for 3D shape has been investigated as well [Osada et al., 2002].

## 4.3 Object Detection

AG assumes that a comparison between two images can be performed only if some objects have been recognized in both sample and target. Similar objects should have similar distribution of geometric features, such as *centroids*, which have to be robust to transformations and slight changes in perspective. Applications like games or educational tools for kids can take advantage of image recognition algorithms that don't rely only on accuracy and generality in detect an object, but also on *similarity* classification.

The concept of similarity is ambiguous: *when two objects can be considered similar?*. We will see how AG allows to set a threshold in order to limit the amount of false positives.

Furthermore, AG aims to recognize objects without any previous training process, limiting the learning step to a single template sampling. Of course this constraint let some challenges arise:

1. Object coverage: the need of recognition a wide variety of different objects, usually classifiable in categories.
2. Intra-category shape variation: the system has to take in account the diversity of shapes that can represent the same object.
3. Inter-category shape ambiguity: shapes that represent different objects can be very similar the system can be "cheated" and generate false positives (or negatives).

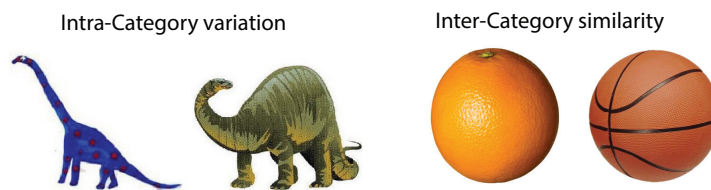


Figure 4.2: AG has been designed to give a solutions to two main challenges: recognize shapes that represent the same object and distinct between shapes that are not similar enough.

To do so, AG extracts image contours from a sample image and then compare the acquired data structure to that one extracted from the captured image in real-time, as showed in Fig. 4.3

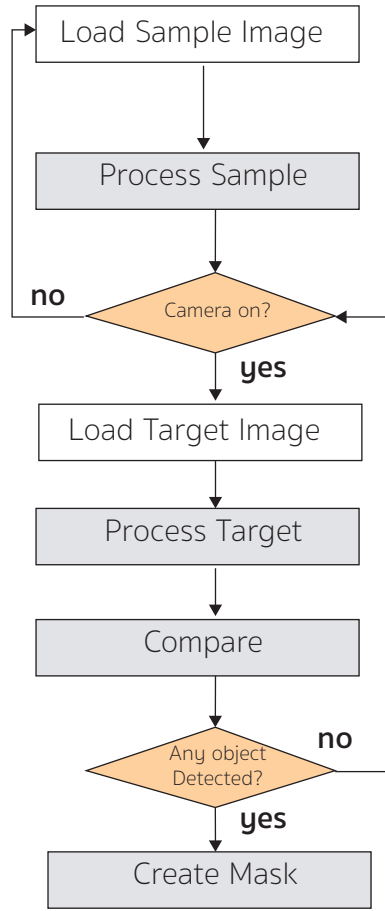


Figure 4.3: AG works as follows:

### 4.3.1 Pre-process image

An image is taken (generally from the camera) and divided by means of the *Rule of Thirds*. The *Region of Interest*(RoI) is the central part of it and it is supposed to be the area where the object to find is centered. Our RoI intuitive considerations have been confirmed by experimental results, as in Fig. 4.4.

The first step is the palette conversion from color to grayscale. In fact color information is not taken in account for contour extraction and thus a grayscale image (one channel) can be processed easily in the subsequent *Image Filtering* operations. For each pixel location  $(x, y)$  in a source image of size  $m \times n$  pixels, its  $k^2 - 1, k \leq m, n$  neighbors are used to compute the resultant pixel at the same location and they are weighted in a way that represents the *kernel* of the filtering operation. The resulting image has the

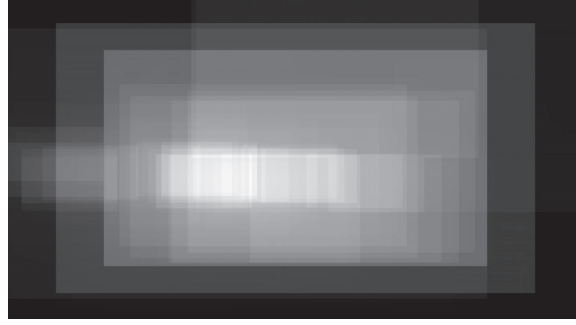


Figure 4.4: RoI in a video streaming analysis [Quang Minh Khiem et al., 2010]. Brighter areas are that ones mostly observed.

same size of the source image and it is given by the convolution of the source with the kernel:

$$dst(x, y) = k * src(x, y) \quad (4.4)$$

Among the filters, we are interested in that ones that can reduce noise. A first option is the *Gaussian Blur*, a low-pass filter that applies a 2D Gaussian smoothing kernel such as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.5)$$

An example is seen in Fig. 4.5.



Figure 4.5: Application of a gaussian filter with  $\sigma = 4$ .

The second option is given by a couple of morphological operations: *Dilation* and *Erosion*. *Morphology* is the common term for operations that process images based on shapes. Morphological operations apply a *structuring element* to an input image  $m \times n$ , creating a destination image of the same

size. The value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with a subset of its neighbors, selected by a specific shape. *Dilation* adds pixels to the boundaries of objects in an image, while *Erosion* removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. In Erosion each output pixel get the minimum value of the pixels inside the kernel. It follows the reduction of intensity of the high value pixels (closer to white color) and the output image tends to be darker. In Dilation it happens the opposite, with each output pixel that get the maximum values among the neighbors, and the resulting image tends to be brighter.

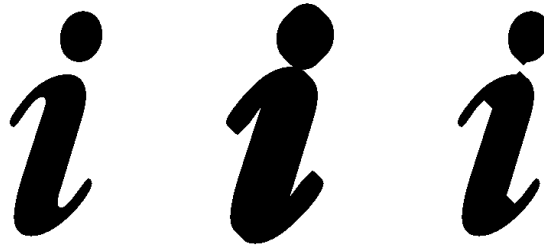


Figure 4.6: Erosion and Dilation operators applied to a shape.

For a further reduction of the unnecessary image data, a *Binary Threshold* operation is applied. If the pixel in a grayscale image are in the range 0-255, the thresholding operation can be expressed as:

$$dst(x, y) = \begin{cases} 255 & \text{if } src(x, y) > threshold \\ 0 & \text{otherwise} \end{cases} \quad (4.6)$$

Thus, each pixel can get only two possible values: 0 and 255 (black or white) Fig. 4.7. Binary Threshold step shows clearly one fundamental request for a suitable object detection with AG: *Object and background have to be distinguishable*. *Shadow Removal* is a fundamental image processing task that has been investigated extensively [Finlayson et al., 2006]. From RGB color space analysis [Baba and Asada, 2003] to algorithms optimized by using SUSAN operators [Zhang et al., 2014], several methods have been



Figure 4.7: Application of a Binary Threshold.

proposed as solutions that aim to recognize in the image the changes in pixels due to the illumination. In AG the problem of removing shadows has been partially solved with the opening operation, that allows to reduce the intensity of the pixels covered by shadows. For better results, a variable threshold approach has been adopted.

### 4.3.2 Find contours

Contour Detection is the next step that AG takes. The task is accomplished by OpenCV Suzuki's algorithm implementation [Suzuki and be, 1985].

`findContours` function retrieve contours from the binary image and store them as vectors of points. A hierarchy structure retains information about the image topology. "For each  $i$ -th contour `contours[i]`, the elements `hierarchy[i][0]`, ..., `hierarchy[i][3]` are set to 0-based indices in contours of the next and previous contours at the same hierarchical level, the first child contour and the parent contour, respectively. If for the contour  $i$  there are no next, previous, parent, or nested contours, the corresponding elements of `hierarchy[i]` will be negative" [Bradski, 2000].



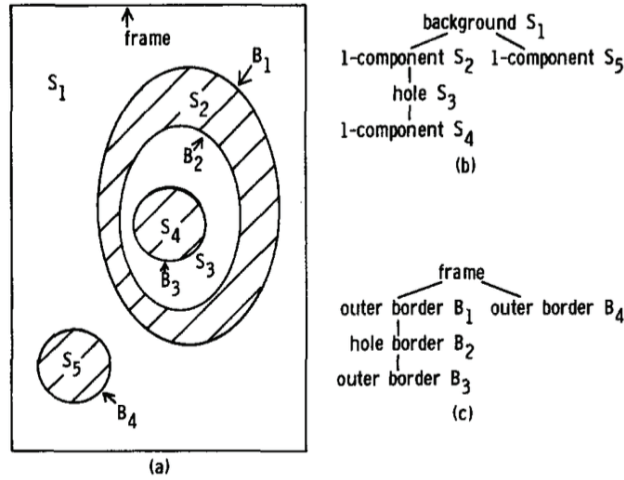


Figure 4.8: A scheme of the topological structural analysis in Suzuki's algorithm.

`findContours` function allows to organize contours hierarchically or not. The two options yield in very different results, according to how much the object is centered in the frame. The number of contours given by Suzuki's algorithm is highly variable, among the factors that influence it, frame size is critical. We will see better the weight of each factor.

### 4.3.3 Process Contours

Once the contours have been found, a two-steps approximation phase is needed to reduce the redundancy of keypoints or the presence of outliers:

1. *Convex Hull*. The convex hull of a set of points  $S$  is the intersection of all convex sets that contain  $S$ . For  $N$  points  $p_1, \dots, p_N$ , the convex hull  $C$  is then given by the following expression:

$$C \equiv \left\{ \sum_{j=1}^N \lambda_j p_j : \lambda_j \geq 0 \text{ for all } j \text{ and } \sum_{j=1}^N \lambda_j = 1 \right\} \quad (4.7)$$

OpenCV implements Sklansky's algorithm [Sklansky, 1982], that has a complexity of  $O(N \log N)$ , it is particularly designed for 2D polygons and is simpler than other Convex Hull algorithms such as *Gift wrapping* [Jarvis, 1973], *Graham* [Graham, 1972] and *Chan* [Chan, 1996].

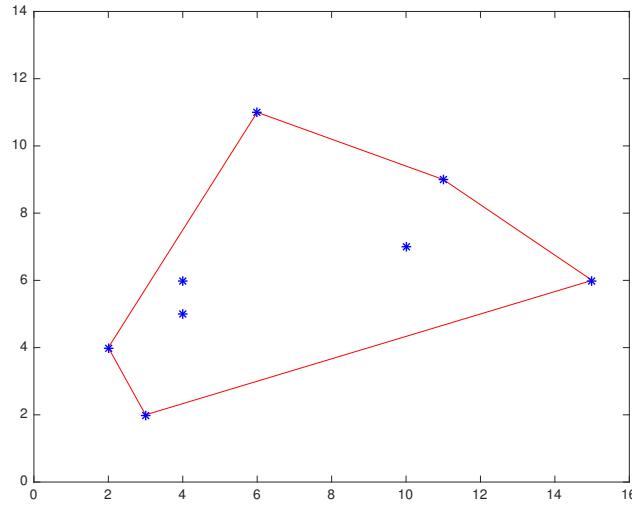


Figure 4.9: Convex hull of a set of points.

2. *Ramer-Douglas-Peucker*. Given a curve with several points (and thus line segments) we need to approximate it with a fewer points set, according to a maximum distance between the original curve and the simplified one. Ramer-Douglas-Peucker (RDP) algorithm uses the *Hausdorff Distance*<sup>1</sup> to generate a subset of points taken from the original curve (see Fig.4.10) [Ramer, 1972] [Douglas and Peucker, 1973].

---

<sup>1</sup>Basically, the Hausdorff Distance is the greatest of all the distances from a point in one set to the closest point in the other set:

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$$

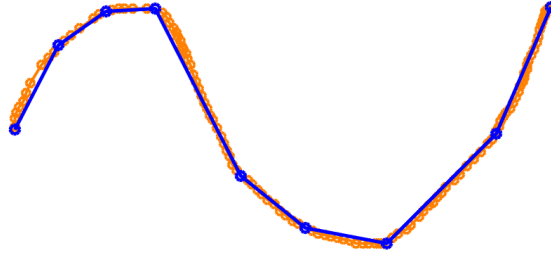


Figure 4.10: A curve approximated to a 9-points set by the RDP algorithm. In most realistic cases RDP complexity is  $O(n \log n)$  but its worst-case complexity is  $O(n^2)$ .

#### 4.3.4 Keypoints Extraction

AG performances are strictly related to how the contours are recognized. From complex objects, made by elaborated textures for instance, we can expect much more contours than simpler ones such as, a flat, well designed, smartphone. In AG we focused on the recognition of three types of objects:

1. *Simple Real*. That is, objects with flat colors and simple shapes.
2. *Complex Real*. That ones with textured colors and multi-shaped geometry.
3. *Hand Drawn*. Objects that are already hand drawn as contours, but they need to be converted in a set of points.

We will see how the three types of objects will give different results. To maximize the object recognition process, two modes have been implemented:

- i. *Monocontour*. Recognition task is based on the external contour of the object. Typical use is for objects taken from pictures and video streaming;
- ii. *Multicontour*. Recognition takes in account also the internal contours of the object. It is ideal for hand drawn objects.

Feature Descriptors are vectors in image recognition task that allow to organize the information that identify an image. Because in SURF and SIFT their construction was computationally demanding, other descriptors showed better performance: BRIEF [Calonder et al., 2010] used simple binary tests between pixels in a smoothed image patch but was sensitive to in-plane rotation, issue solved by ORB [Rublee et al., 2011]. AKAZE [Alcantarilla et al., 2013] exploits the benefits of nonlinear scale spaces. In AG the function of feature descriptors is fulfilled by the *centroids*, that are extracted from the contour as follows:

---

**Algorithm 2** AG Centroids extraction

---

```

1: CalculateCentroid(Cntr)
2: Cntr1  $\leftarrow$  BisectHorizontally(Cntr) + BisectVertically(Cntr)
3: for eachCntr1inCntr1 do
4:   Cntr2  $\leftarrow$  BisectHorizontally(Cntr1) + BisectVertically(Cntr1)
5:   for eachCntr2inCntr2 do
6:     CalculateCentroid(Cntr2)
7:   end for
8: end for

```

---

The centroids are distributed mostly in the contour, as in Fig. 4.11:

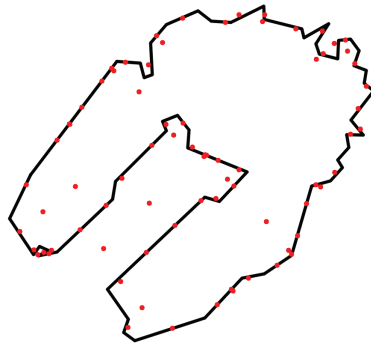


Figure 4.11: An object and its centroids.

### 4.3.5 Shape Matching

The comparison between the sample object and the frame object is based on the *Cosine Similarity* (CS) measure<sup>2</sup>. CS measures the cosine of the angle between two vectors. As:

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta \quad (4.8)$$

it yields:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (4.9)$$

where in our case centroids are the components  $a_i$  and  $b_i$ . In AG we introduced a threshold that can limit the presence of false positives. We observed that for high values ( $> 95\%$ ) the algorithm detects the sample object with high accuracy, but limited in terms of orientation, scale and position. For lower values, the algorithm is more robust to object transformation and can compare positively also object that are similar. Shape similarity gives interesting results when sample objects are compared applied to hand drawn objects, as we see in Fig. 4.12, moment invariant properties allow to match different representations of the same object.



Figure 4.12: An object can be drawn and then compared positively with the original one.

### 4.3.6 Create Mask

A matching object is returned by the `findObjects` method in two ways:

1. with all the original image;

---

<sup>2</sup>Cosine Similarity measure is largely used in *text mining*: a document is characterized by a vector where the value of each dimension represents the number of times a term appears in the document. Cosine similarity then gives the “distance” between two documents, as a measure of similarity between their subject.

2. masked with an alpha channel, as in Fig. 4.13.



Figure 4.13: Object extraction by masking.

The diagram of flux in Fig.4.14 lists the basic steps of AG pre-processing and comparison, while Fig. 4.15 shows how an object can be “extracted” from reality to be used in virtual environments.

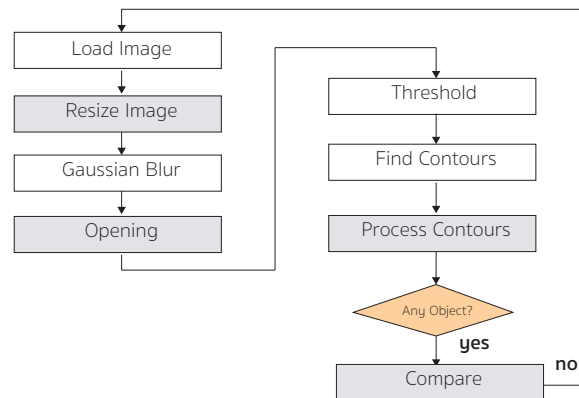


Figure 4.14: The operations necessary for processing each image and compare it with the sample.

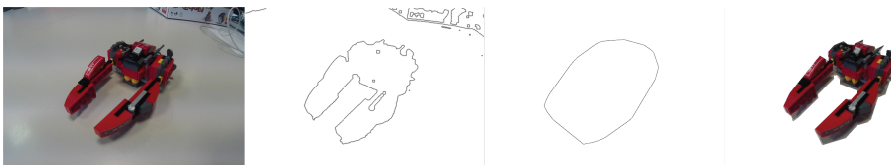


Figure 4.15: The sequence of operations for extracting an object.

## 4.4 Implementation

AG has been implemented in mobile demos targeted for children in pre-school and school age. Object recognition ability continues to develop through the years, emerging in infancy, when kids perceive simple shapes to childhood, when neuroimaging research revealed the first traits of an adult ability [Nishimura and Behrmann, 2009].

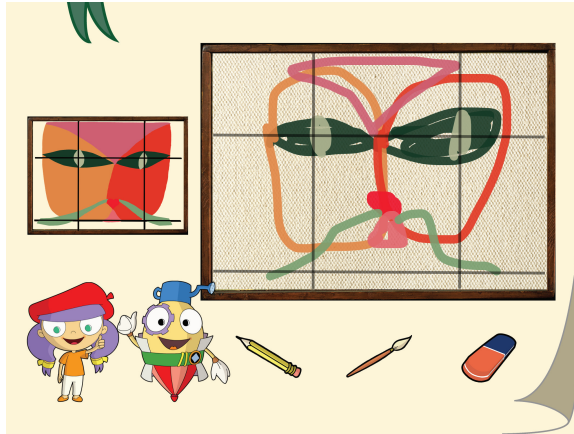


Figure 4.16: A mobile game for teaching art to kids by drawing the works of famous artists. In this image, “Cat and Bird” by Paul Klee (1928) can be reproduced by means of a series of levels where kids have to draw details of the painting. If their drawing looks enough similar to the original, they jump to the next level.



Figure 4.17: Moments invariants allow to compare objects under rotation, scale, translation and slight perspective transformations, while shape matching is enough robust to match different representation of the same object. In the example, a smartphone is detected in different orientation and also when is hand drawn.

## 4.5 Evaluation

AG has been tested by means of an image set partially based on Caltech101 [Fei-Fei et al., 2007] and custom images, in total 215 images where the objects to be recognized was centered. Tests have been made on mobile (iPad 4th Generation with iOS 7-8-9) and desktop systems (i5 and i7 Intel processors). No particular parallelism optimization has been given to speed up the AG algorithm. Embedded cameras (iOS) or external web cams (PC) worked at a frame rate between 10 and 29 fps. Evaluation of AG has been done according two different benchmarks:

1. *Intra-Category*, that is, the AG performances within the same category of objects
2. *Inter-Category*, how AG performs comparing sample and frame images taken from different categories.

Given the following measures:

$$\begin{aligned}
 Accuracy &= 100 \times \frac{TruePositives + TrueNegatives}{TotalCompared} \\
 Precision &= 100 \times \frac{TruePositives}{TruePositives + FalsePositives} \\
 PositiveRate &= 100 \times \frac{TruePositives}{TruePositives + FalsePositives} \\
 RecognitionRate &= \frac{Precision \times PositiveRate}{100}
 \end{aligned} \tag{4.10}$$

In Intra-Category testing, Positive Rate is far from the results that can be achieved with classification methods. However, we compared the *Recognition Rate* and the *Positive Rate* as a measure of how often objects found in the target images are wrongly recognized as similar to the sample object. As you can see from Fig. 4.18, the contribution of false positives is really low, except for Binocular category.



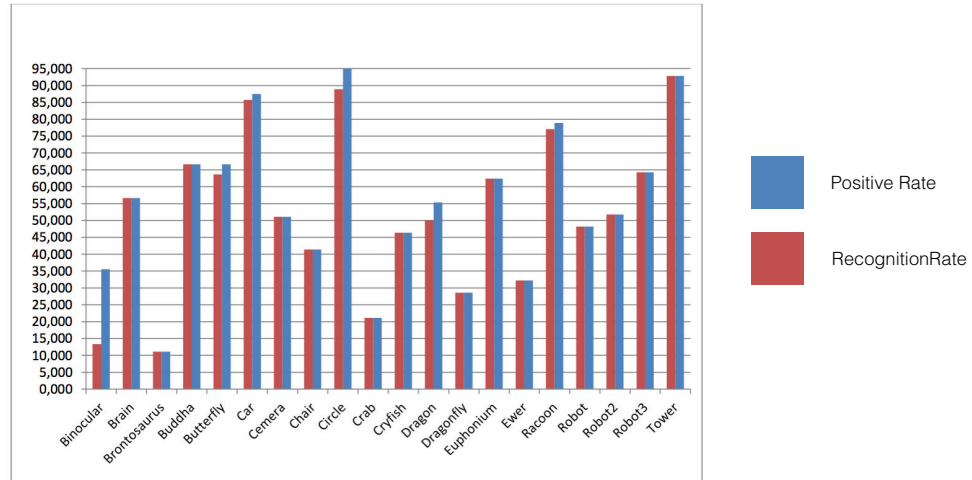


Figure 4.18: Intra-category results for some categories taken in account.

In Inter-Category evaluation, we were interested to measure how often an object of a category can be “confused” with another category, so we introduced the *True Negative Rate*(TNR) as follows:

$$TrueNegativeRate = 100 \times \frac{TrueNegatives}{TrueNegatives + FalsePositives} \quad (4.11)$$

A low TNR means a high probability to be classified as a similar object, that is, they are “misunderstood”. From Fig. 4.19 we see that there are categories of objects that can be wrongly classified with a probability of more than 30%. Such result, that we interpreted as a measure of the similarity between two objects that belong to different categories, it is something we are interested to exploit.

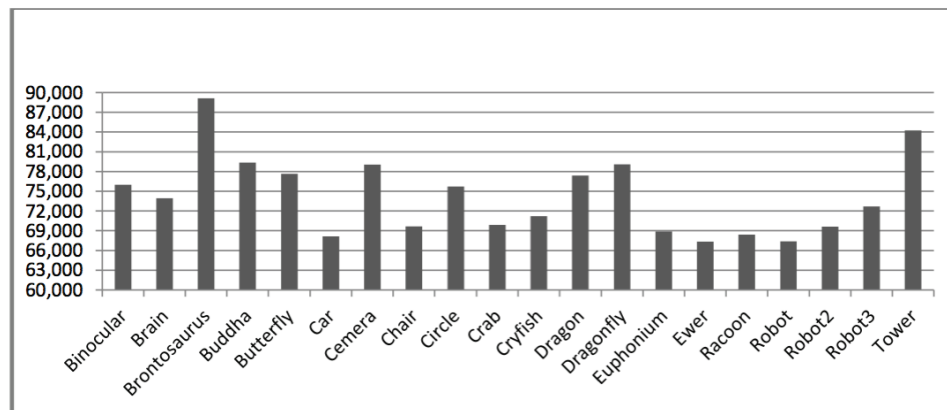


Figure 4.19: Inter-category results for some categories taken in account.

### 4.5.1 Discussion

The evaluation made on a first set of images showed that the presence of false positives is really low in intra-category, as we can see from Fig. 4.18. The blue and red bars are almost the same. An exception is the binocular object that is correctly recognized only for the 13,33%. It means that when something is identified as an object, with its contour developed around the center, almost certainly will be recognized positively. For the inter-category test, we found a 25.76% of probability on average to recognize an object as being part of the wrong category. This result, although it seems elevated, it is what we exploit as a feature for extend object recognition to similar objects that belong to different categories.

It is important to remember that AG is not tracking images. At the first stage, the algorithm results are still affected by camera noise and there are no descriptors as in SIFT, ORB, BREAK, etc. for tracking the object by means of the image features. The peculiarity of this approach doesn't permit to compare AG with other tracking systems, like PTAM [Klein and Murray, 2009].

## 4.6 Future Work

AG was designed with mobile devices in mind, as an alternative to more sophisticated object recognition algorithms that need a training set. Deep Neural Networks perform better than any other in image classification but the training process need an accurate choice of the proper features to create the model and are not immune to side effects [Nguyen et al., 2014], as seen in Fig. 4.20.

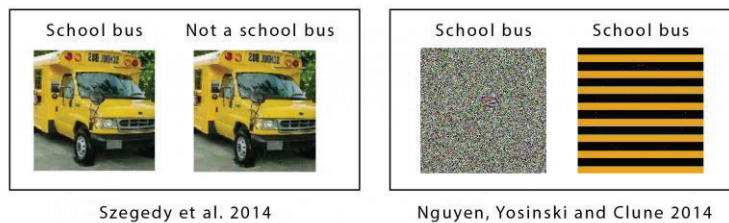


Figure 4.20: Two examples of a “fooled” Deep Neural Network. *On the left, a couple of school buses:* the second image has been modified on a detail, imperceptible to human eye, but significant enough for the net to don't classify it as a “school bus”. *On the right:* two abstract images that are classified with high confidence as “school bus”.

The AG approach is different, it considers the “similarity” between different objects as a result that we can use exploit as well, in some way. For this reason the tests we made had to take in account the ambiguity of the concept of similarity. AG looks suitable for applications that doesn’t need the highest accuracy on the market but a quick setup. As AG needs just one sample and doesn’t use any training set, it is particularly suitable for mobile devices and has encountered the interest of gaming and educational community that consider kids as a an appropriate audience, and AG as a creative support for their exploration of shapes, colors and geometries similarities. AG finds its ideal continuation in 3 dimensions. How moments invariants and contours analysis can be extended to 3d is the next challenge.

## Chapter 5

# Hand Motion in Virtual Reality: GLOVR

Virtual Reality (VR) and Augmented Reality (AR) are deeply changing the game industry, introducing new engaging ways to interact and play. The virtual worlds are, in turn, celebrating a new age for computer graphics and storytelling. Furthermore, VR and AR are finding important applications outside the world of gaming and they look promising in movies, training, marketing and advertising. Consequently a new generation of hardware and software tools for both VR and AR have entered to marketplace, while many others are constantly under development [Lee et al., 2015].

However, despite the rapid technical progress in headset, display and head tracking [LaValle et al., 2014], the navigation in virtual worlds remains problematic. Mouse, joysticks and keyboards are still the most popular control devices, but they constrain the user to the desk area. Gamepads provide more freedom, but they severely limit the interaction with the virtual objects. Furthermore, all these devices gives occasionally rise to counter-intuitive movements of the user in the virtual environment, producing very fastidious reactions, when not real sickness.

In addition, no reliable solution has yet been proposed to allow interacting with the real world during a VR session without pausing or stopping the application. For instance, while at present the average playing time in VR is still limited, mainly because of the motion sickness, future applications will last much longer, with the need of tools allowing fast interaction with not VR applications during VR sessions. GLOVR has been designed in front of these challenges, as a hand controller that allow users to play and communicate within VR environments in a very natural way, namely using gestures, distinguished in *classified* and *instant*, and natural language. We will see how GLOVR works in more details.

## 5.1 Related Work

The huge success of 3D user interface technologies in game industry since Nintendo Wii has revolutionized how people play, spreading the game culture to a wider audience made of “casual” gamers [Verplaetse, 1996]. The main reason has to be attributed to the introduction of a natural interaction interface, based on the inertial sensors integrated in the game controller (*Wiimote*). Since *Wiimote*, every control system took into account the importance of an easy-to-use interaction, which can let people reproduce the actions they do in reality (singing, hitting objects, running, etc.). Several studies have been made about the benefits of these systems (gesture controls, stereoscopic 3D, head trackers), either considering the user interfaces in isolation or taking in account the mutual influence in the gaming experience [Kulshreshth and LaViola, 2015].

Hand gesture recognition is an important topic of the field of natural interfaces, that has been proposed in several contexts, from education to rehabilitation, from games to live music performances.

According with the most popular approaches, while tracking systems made with depth sensors can process frames and detect three-dimensional hand shapes in the environment, inertial and bend sensors track accurate movements in local coordinates [Sharp et al., 2015, Sutton, 2013]. New radar-based technologies, like Google’s *Project Soli*, look promising.

In both solutions the hand and/or fingers positions are recorded in a small interval of frames and classified to predict the most probable hand gestures. The most used classification methods (Hidden Markov Models, Neural Networks), can discriminate between static and dynamic gesture recognition [Rabiner, 1989, Xu et al., 2012, Hasan and Abdul-Kareem, 2014].

The success of wearable devices has interested different areas, from sports and wellness to gaming. The availability of accurate inertial sensors and the intensive research on pattern recognition methods have given the conditions for developers to create applications that track personal data and promise to improve the lifestyle and the daily habits of users. Despite the popularity of such devices, in the game industry the diffusion of wearable controllers is still not commercially significant. However, a visible interest has always been present in the intersection between games and personal training, rehabilitation or education [Xing et al., 2009, Vasudevan et al., 2015]. The idea of a hand glove as a game controller is not new, of course, and has been considered a fascinating technology since the 80’s [Zimmerman et al., 1987]. However,

so far the wearable hand solutions have not encountered the acclaim of a huge audience, although the appealing design often reminds science fiction movies and cyberpunk imaginary.

## 5.2 Overview

GLOVR is composed by a PCB hosting:

1. a microcontroller;
2. a 9-axis Inertial Measurement Unit (IMU) with accelerometer, gyroscope and magnetometer sensors;
3. an analogue microphone.

and connected to a PC via Bluetooth or USB (Fig. 5.1).

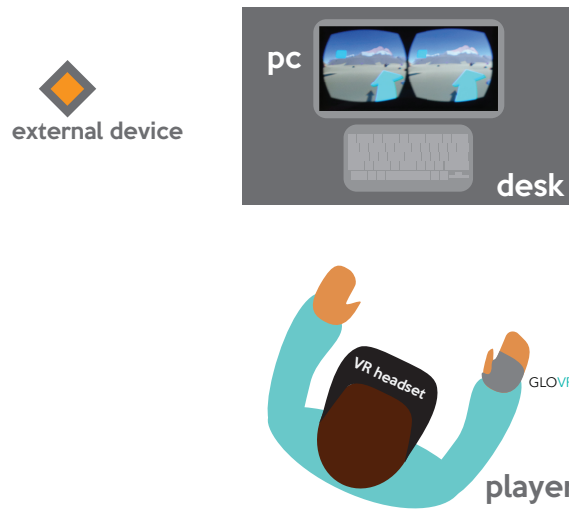


Figure 5.1: A GLOVR setup.

The system is packaged into a small box attached to a textile glove that can be worn indistinctly on both right and left hands (see Figure 5.2).

In order to guarantee high usability, the fingers have been excluded from gesture recognition (also making the glove much more practical to wear for longer sessions). On the other hand, capturing only palm movements has been assumed sufficient for satisfactory interactions with most applications. GLOVR communicates with the pc via Bluetooth. Once the raw data have



Figure 5.2: The GLOVR device.

been preprocessed by a *Data Handler*, it sends its results to an *Action Manager* that translates classified gestures and instant controls in game actions and runs the speech recognition service (see Figure 5.3) according to the messages the controller receives from the microphone. The libraries developed for GLOVR have been implemented in Unity as a *package*, and so is also our 3D User Interface, that shows the current state of the hand controller.

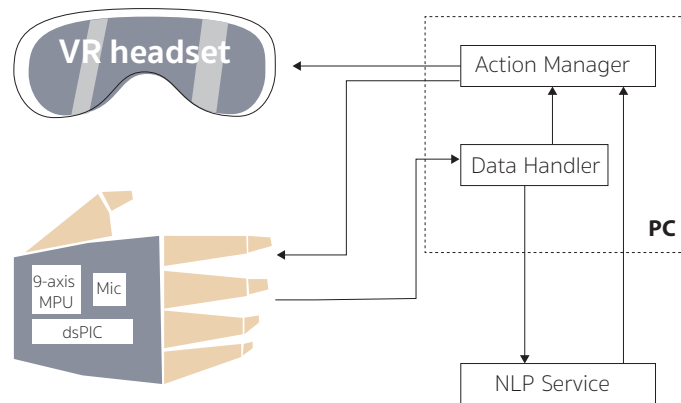


Figure 5.3: The GLOVR architecture: all the data are sent to a *Data Handler* that detects the instant controls and classifies the hand gestures. If a *Mic Activation* is triggered, it send a request to the *Natural Language Processing Service* connected to Wit.ai. All the directions and classified hand gestures are communicated to the *Action Manager* that translate them in character playing actions.

### 5.2.1 Design Architecture

Let us see GLOVR design with more detail. Data Handler receives pre-processed packets from the GLOVR firmware and distinguish them in voice sequences to send to the NLP service and movement data for the Action Manager (Fig. 5.4).

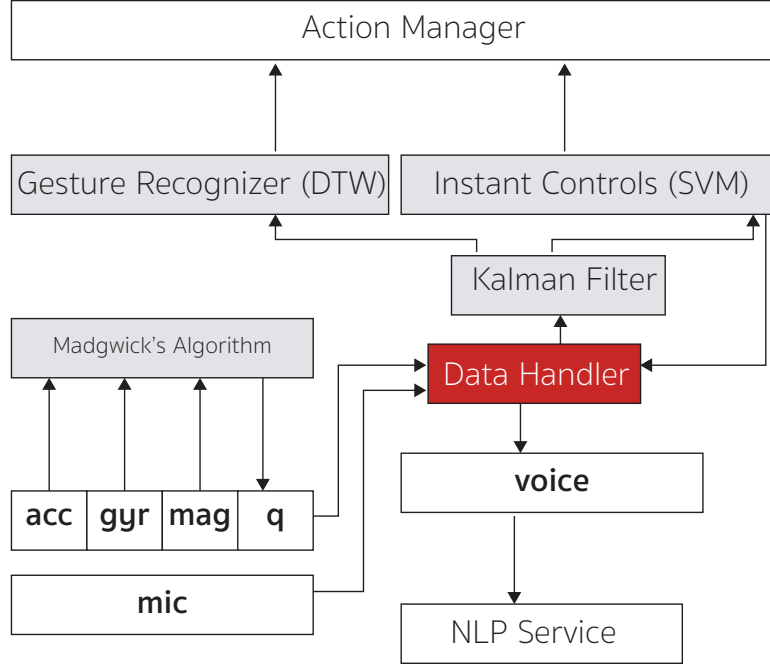


Figure 5.4: Data Handler controls the flux of data between hardware and software.

#### Directional and Instant Controls

As any common input device is a combination of buttons, levers, touch pads and wheels and offers a universally set of actions and directions, our solution had to efficiently reply them as a hand *buttonless* controller. In games, while directions are usually mapped in a continuous range of values by means of directional keys or digital/analogic sticks, actions are triggered by a combination of pressed button and directions. Thus for GLOVR it was necessary to map all the directions in the 3D sphere of the rotations calculated by the inertial sensors raw data. We opted for a supervised SVM implementation [Taranta II et al., 2015], that separates correctly the hand rotations reserved to the directions with that ones deputed to instant gestures. Classifi-



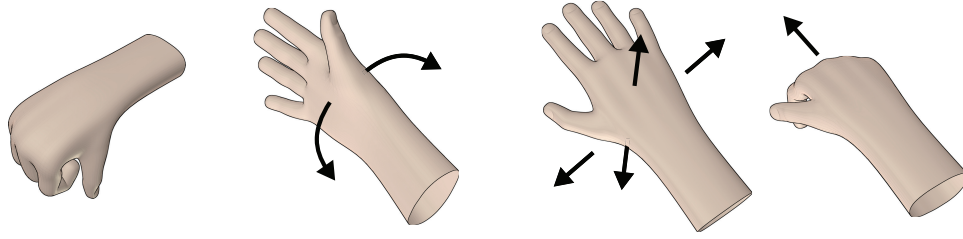


Figure 5.5: Some instant gestures.

cation tasks for instant directions and gestures, as well as gesture recognition, run into the Data Handler (DH) software module. Its output, affected by error generated by inertial sensors and classification inaccuracy, need to be filtered in some way to reduce discontinuity in motion. For such reason a *Kalman filter* [Kalman, 1960] has been implemented, which generates smooth transitions between all the possible gestures and directions, a necessary condition for navigating naturally in VR <sup>1</sup>. The filtered data are then sent to the Action Manager (AM), a module strongly integrated with the host programming environment. For instance, in Unity AM is part of the C# scripts that create the game. According with the data received, AM generates the proper actions for the player's character (see Fig. 5.6).

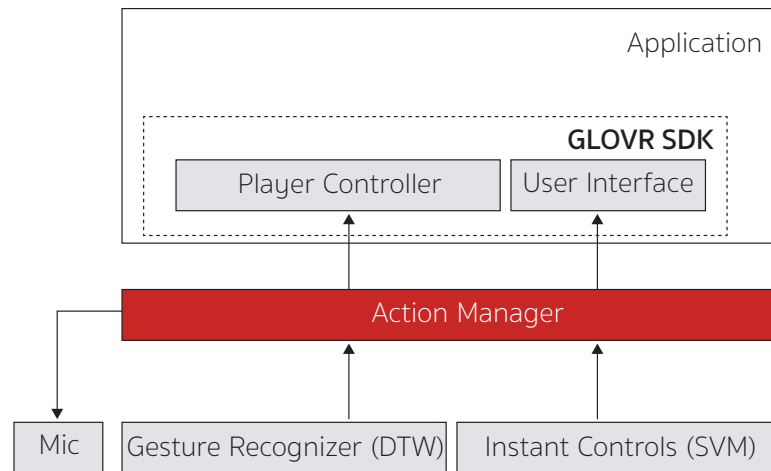


Figure 5.6:

<sup>1</sup>Kalman filter is an iterative two-steps algorithm that works as follows: starting from a linear system with Gaussian errors, at first step it predicts a future state and once it receives data, in the second step it corrects its prediction by means of the current state.

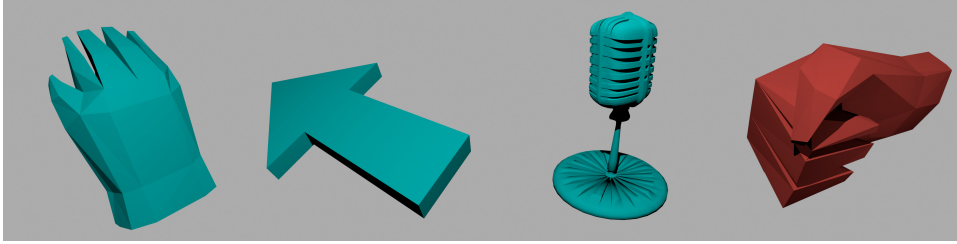


Figure 5.7: The four 3d icons modeled for the User Interface. *Top-Left*: hand model that appears when a gesture is being classified. *Top-Right*: arrow that is visible when getting directions. *Bottom-Left*: the microphone model for Natural Language System activation. *Bottom-Right*: the punch for hitting the enemies.

Thus the *Instant Controls* are grouped in two: *directions* and *actions* <sup>2</sup>:

- *Backward/Forward*. A vertical hand orientation moves backward/forward with a speed dependent on the angle
- *Left/Right*.
- *Jump*. A smooth “pitch” rotation of the palm is translated in a jump.
- *Mic Activation*. Microphone is activated with the palm oriented horizontally and kept close to the mouth giving a sort of “James Bond” style position.
- *Hit*. An action is performed with a fast “punch” forward movement.

In GLOVR the instant gesture icon appears on bottom right of the VR view as long as it was considered a suitable location on the Oculus screen view. The correct implementation of a 2D/3D GUI in VR is still an open research topic [Silva et al., 2014].

### Gesture Recognition

GLOVR represents a simple and practical solution suitable for intuitive and easy interaction with VR environment, that classifies static and dynamic gestures in real-time. To achieve this, gesture recognition task in GLOVR needed an algorithm able to handle 3D hand inputs of various length, that

---

<sup>2</sup>we will see how *actions* can be also performed by gesture recognition task, it is up to the game designer to design how to map them.

may be constituted by a sequence of movements that occur over a variable interval of time. To this purpose we adopted a *N-Dimensional Dynamic Time Warping* (ND-DTW) method that can compute similarities through the concept of *distance* between two  $N$ -dimensional time-series [Gillian et al., 2011]. DTW has been applied to several different fields, from database indexing [Ding et al., 2008] to handwritten recognition [Vuori et al., 2001], gesture recognition [Héloir et al., 2006] and speech recognition [Vlachos et al., 2003]. DTW in one-dimension works as follows. Given two time-series,  $\mathbf{x} = \{x_1, x_2, \dots, x_{|\mathbf{x}|}\}$  and  $\mathbf{y} = \{y_1, y_2, \dots, y_{|\mathbf{y}|}\}$  with lengths  $|\mathbf{x}|$  and  $|\mathbf{y}|$ , construct a *warping path*  $\mathbf{w} = \{w_1, w_2, \dots, w_{|\mathbf{w}|}\}$  so that:

$$\max\{|\mathbf{x}|, |\mathbf{y}|\} \leq |\mathbf{w}| < |\mathbf{x}| + |\mathbf{y}| \quad (5.1)$$

and the  $k$ th value of  $\mathbf{w}$  is given by  $\mathbf{w}_k = (\mathbf{x}_i, \mathbf{y}_j)$ . The minimum total warping path is that one that minimizes the *cost matrix*  $\mathbf{C}$ . Given:

$$DIST(i, j) = \sqrt{\sum_{n=1}^N (i_n - j_n)^2} \quad (5.2)$$

for  $N$ -dimensional time-series, the total distance across all  $N$  dimensions is used to construct the cost matrix  $\mathbf{C}$ . Training is made by creating a template for each gesture that is going to be classified. In a class, templates are the training samples that minimize the normalized total warping distance against the other ones. Accordingly, a  $N$ -dimensional time-series  $\mathbf{X}$  can be classified as " $k$ th gesture" if, among all the templates, it minimizes the normalized total warping distance between  $\mathbf{X}$  and the  $k$ th template.

A critical aspect of ND-DTW is to set a proper threshold for each gesture recognition, in order to reduce the risk of false positives during classification. The problem, better known as *gesture spotting* is faced setting up the classification threshold for each template as the average total normalized warping distance between all the couple of training examples, augmented by their standard deviations.

The main advantages of this method are:

1. it works well for time-series of different length (see Fig. 5.8):

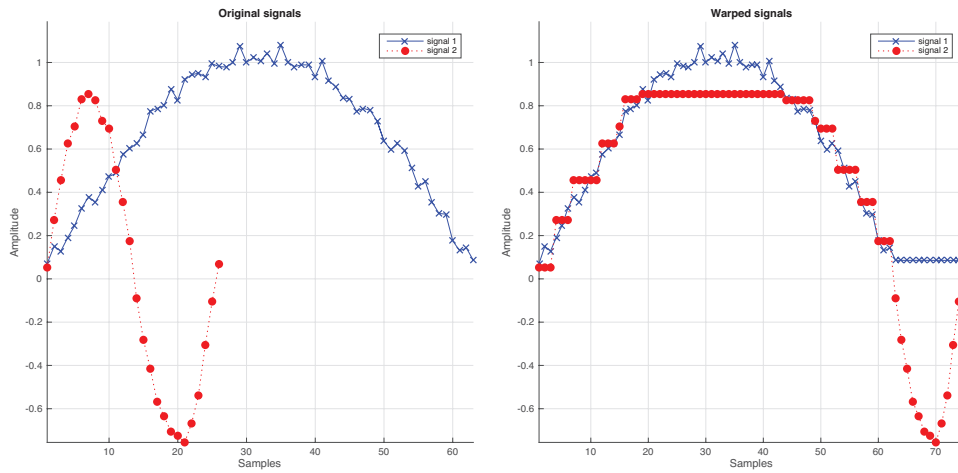


Figure 5.8: The aligned input signals. The red original signal on the left have been “warped” to fit the geometry and the length of the blue one.

2. it provides a measure of the distance between the data to be compared, as in Fig. 5.9:

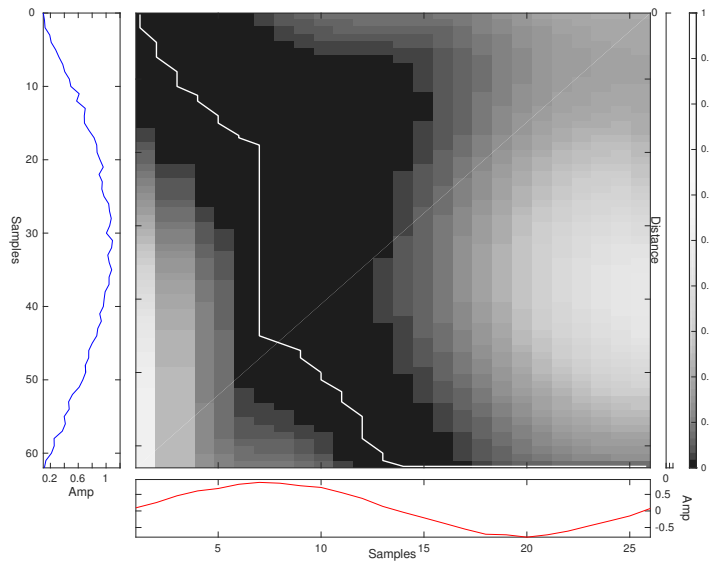


Figure 5.9: The optimal path calculated by the DTW (the white line). The goal of DTW is to minimize the distance between two signals.

3. it guarantees better accuracy with a smaller training set, in comparison

with other methods such as Hidden Markov Models [Carmona and Climent, 2012].

As for gesture classification, a set of 4 gestures (see Fig.5.10) has been designed and tested in a variety of circumstances within the demos, each one trained by a limited number (20) of 6-dimensional samples. As proven by the previously mentioned research, the minimum number of training examples that gives a minimum of 90% of accuracy is 12, thus we considered our setup robust enough. A training phase generates a template and a threshold for each gesture. Then, in operation, the distance between the acquired data and each threshold is computed in real time to distinguish gestures corresponding to the models from all the others, successfully solving the gesture spotting problem. Along with such a recognition, GLOVR calculates also a *Gesture Strength* (GS), suitable for applications where the intensity of a gesture has to be taken in account (such as, for instance, in the case of an action game where the player can hit an enemy with different levels of power).



Figure 5.10: The four gestures under test for GLOVR and their hand movements.

### Natural Language System

The more evident limits in natural language interaction are clearly in the relationship between what a user says and the consequent action taken by the computer. The difficulties in achieving this task can be essentially grouped in three types:

- Speech recognition is compromised not only by audio quality but also by wide differences (by age, sex, nationality, etc.) in voice and pronunciation. Thus it is extremely hard to create a universal model that cover all the targets.
- Understanding the user's intent is a complex task that has to take in account the intrinsic ambiguity of human language, that is, the same utterance can have a different meaning for different users.

- All natural language processing tasks need generous processing resources, in practice still available only offline. This implies that on-line voice assistant systems are subject to trade-off due to latency and bandwidth.

However, the progress in speech recognition and natural language processing have moved tech companies and researchers of future opportunities and applications. The success of personal voice assistants such as Apple’s Siri, Microsoft’s Cortana or Google Now for scheduling, traffic navigation or other simple management tasks is a clear indication of high interest in developing “speech-controlled” systems. An increasing number of free-to-use or licensed natural language services have been created, and now it is possible to develop applications embedding a speech recognition system in several languages [API.ai, ].

In games, the exploitation of this type of interaction is still at an early stage and it has already been commercialized in educational games for kids [ToyTalk, ]. Other wearable solutions have been experimented in rehabilitation, exploiting the potential advantage of the combination of motion capture and speech recognition [Pereira et al., 2011]. Also, natural language interaction looks promising particularly for use as a complement of other solutions in order to increase functionality [Cheong et al., 2014, Zhu et al., 2014].

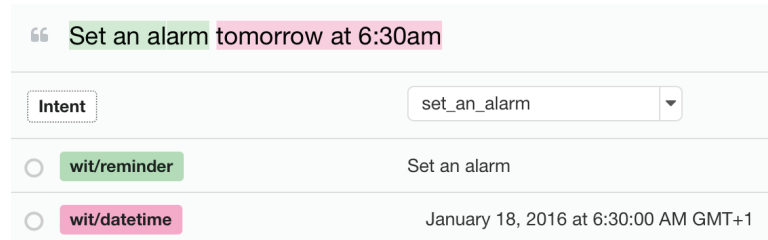
In GLOVR the microphone has been implemented in the palms and it’s actionable by means of a simple gesture. Although in VR a voice input could be integrated in the headset or in front of the user, we found that a wearable solution could cover several other contexts than VR where GLOVR can work efficiently. Thus in GLOVR we aimed at realizing two kind of operations:

- *Voice-activated tasks.* The user sends voice requests that are recognized as commands for the execution of specific tasks, such as, for example: activation of messenger services, remote regulation of connected devices.
- *Gameplay.* The user communicates inside the game with other characters or gives instructions to the game system (such as: switch of game modes, dialogues with characters).

Speech recognition has been implemented using Wit.ai, a natural language processing service that allows to turn speech and text into actionable data [Wit.ai, 2016]. Wit.ai recognizes user’s commands and retrieves data for the subsequent action after it has been trained. To do we need to define *intents* by means of natural language expressions. Let us give an example:

1. we create a new intent `set_an_alarm` with just two expressions like “Set an alarm tomorrow at 6:30am” and “Wake me up tomorrow at 7am”;

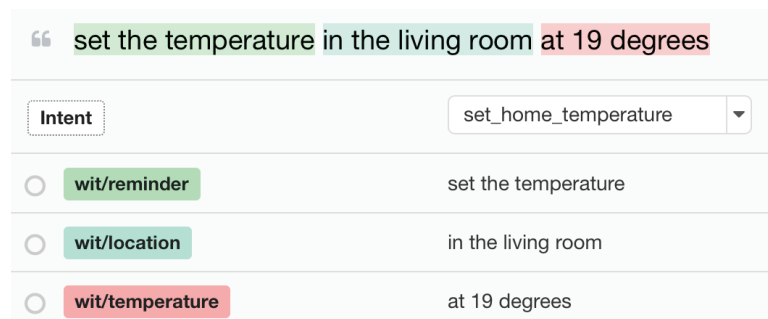
2. we connect *entities* to the related parts of each expression as in Fig.5.11:



“ Set an alarm tomorrow at 6:30am ”	
Intent	set_an_alarm
<input type="radio"/> wit/reminder	Set an alarm
<input type="radio"/> wit/datetime	January 18, 2016 at 6:30:00 AM GMT+1

Figure 5.11: Connected device such as an alarm clock can receive commands via natural language.

We can populate our natural language system with more elaborated intents as `set_home_temperature`, based on expressions like “Set the temperature in the living room at 19 degrees” or “Decrease the temperature in the bedroom to 17 degrees”, and entities as:



“ set the temperature in the living room at 19 degrees ”	
Intent	set_home_temperature
<input type="radio"/> wit/reminder	set the temperature
<input type="radio"/> wit/location	in the living room
<input type="radio"/> wit/temperature	at 19 degrees

Figure 5.12: A heater can be controller at distance by natural language commands which set the temperature in the different rooms

Wit.ai offered a set of built-in entities that cover a wide variety of situations and needs: from `wit/agenda_entry` that extrapolates agenda items from text, to `wit/email` that detects email addresses, from `wit/url` that captures an URL address to `wit/wikipedia_search_query` that send queries to Wikipedia. During a game session, a user’s request like “Say John Smith to call me later” is in relationship with an intent like *Send a message* and the entities *contact* and *message\_body* can be associated in it (see Figure 5.13).

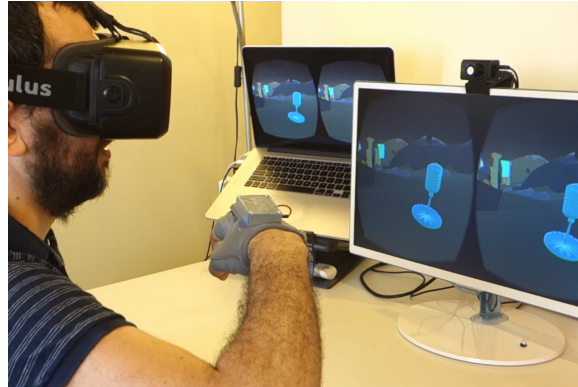


Figure 5.13: *Mic Activation* pose.

A scheme of the natural language implementation is given in Fig. 5.14. The Action Manager receives the intent from the NLP service and send to the *Intent Listener*, which runs via the GLOVR SDK and can connect the client with external applications and devices.

## 5.2.2 Implementation

### System Architecture

GLOVR hardware is composed by: a 16bit Microchip dsPIC33x microcontroller that processes the data, along with the IMU (InvenSense 9250) and the a analogue microphone (Wolfson WM7120). All the circuitry is embedded in a 1.57"x 1.18" PCB hosted in a 3d printed case, on top of a textile glove. The device can transmit via Bluetooth or MicroUSB connection. Power is supplied by a Li-Ion 3.7V 240mAh battery or via MicroUSB. Code runs as an C-style API (for desktop apps and Unity) or a Node.js module (for Javascript). The latter solution looked interesting for the raising diffusion of VR web apps. Both are cross-platform and can run with limited hardware resources.

Serial port connection has been implemented in two ways:

- A Java web server, that provides data communication for desktop and web apps.
- A Unity package, that calls the API functions and allows a easy integration of GLOVR inside a game.



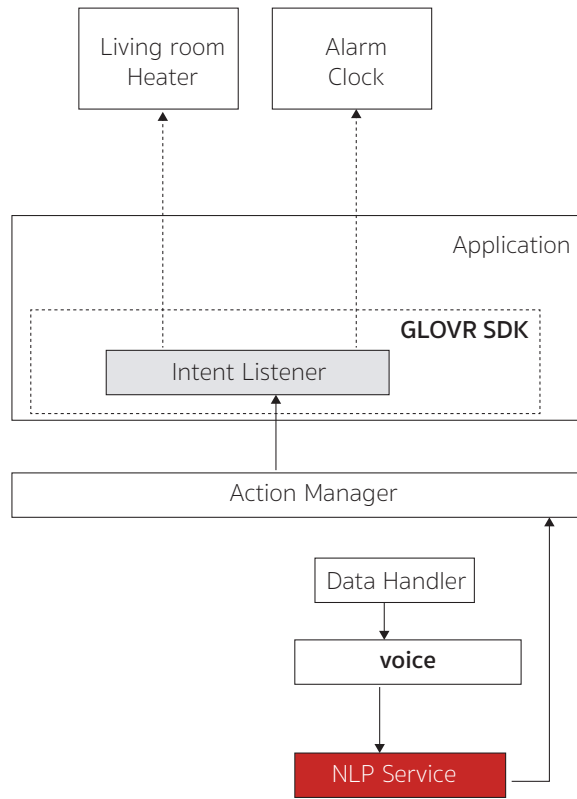


Figure 5.14: Scheme of the natural language interface. In our example configuration, a heater placed in the living room and an alarm clock can be controlled by voice commands sent from GLOVR.

## Hand Movement

In our system, all the data about hand movements are acquired by a InvenSense MPU-9250 9-axis device, featuring accelerometer, gyroscope and compass. The 9-dimensional vector of motion data is acquired with a frequency of 30Hz and pre-processed internally by the MPU by means of a Kalman-based orientation filter developed by Sebastian Madgwick that fuses raw data and returns an orientation quaternion<sup>3</sup>. Madgwick's algorithm gets as constants the sample frequency and a **beta** value, that tunes the response of the algorithm to changes in sensors data. Every loop the `update()` function accepts as input parameters the raw data taken from the accelerome-

<sup>3</sup>Quaternions are members of a noncommutative algebra invented by W. Hamilton. They can be represented as  $q = a + bi + cj + dk$  and briefly, they are preferred over *Euler Angles* because are free of singularity issues and easier to handle in operations.

ter, the gyroscope and the magnetometer. Madgwick's algorithm was preferred over a common Kalman Filter because provide a comparable accuracy and they are less expensive in terms of calculations, letting the MPU go faster [Madgwick, 2010].

---

**Algorithm 3** Madgwick's sensors fusion update() algorithm

---

```

1: g_vec  $\leftarrow$  (gyroX, gyroY, gyroZ)
2: a_vec  $\leftarrow$  (accX, accY, accZ)
3: m_vec  $\leftarrow$  (magX, magY, magZ)
   Rate of change of quaternion from gyroscope
4: qDot =  $0.5 \times f(\mathbf{q}, \mathbf{g\_vec})$ 
   Normalise accelerometer measurement
5: recipNorm = invSqrt(a_vec)
6: a_vec $\times$  = recipNorm Normalise magnetometer measurement
7: recipNorm = invSqrt(m_vec)
8: m_vec $\times$  = recipNorm
   Auxiliary variables to avoid repeated arithmetic
9: process_auxiliary_variables(m_vec, q)
   Reference direction of Earth's magnetic field
10: update_reference(m_vec)
   Gradient decent algorithm corrective step
11: s  $\leftarrow$  process_corrective_step(a_vec, m_vec, q)
12: recipNorm = invSqrt(s)
13: s $\times$  = recipNorm
   Apply feedback step
14: qDot = qDot + s
   Integrate rate of change of quaternion to yield quaternion
15: q+ = qDot  $\times$  (1.0/sampleFreq)
   Normalise quaternion
16: recipNorm = invSqrt(q)
17: q $\times$  = recipNorm

```

---

From them, we calculate hand rotation angles *roll*  $\phi$ , *pitch*  $\theta$  and *yaw*  $\psi$  with a small amount of noise and drift error with the following formulas:

$$\begin{aligned}
\phi &= \text{atan2}(2(q_0q_1 + q_2q_3), 1 - 2(q_1^2 + q_2^2)) \\
\theta &= \arcsin(2(q_0q_2 - q_3q_1)) \\
\psi &= \text{atan2}(2(q_0q_3 + q_1q_2), 1 - 2(q_2^2 + q_3^2))
\end{aligned} \tag{5.3}$$

where  $q_0, q_1, q_2, q_3$  are the quaternions calculated by Madgwick's algorithm.

### Natural Language System

GLOVR acquires audio data thanks to a silicon MEMS analogue microphone, a Wolfson WM7120, that features a high SNR 75dB and a  $140\mu\text{A}$  of supply current, producing a Pulse Density Modulated (PDM) audio stream. The microphone, located on top of the controller, is activated with an intuitive hand gesture and used for voice command. As default, the allowed communication length is 2-3 seconds, that represents a good trade-off between duration and complexity of user's requests. Wit.ai service has been implemented within the Unity package. In 4 the main algorithm is listed and 5 summarizes the key steps in processing the voice sample. Once speech is recorded by microphone, an http request is sent to the Wit.ai service that returns a **data** structure made by the recognized **text** and an array of **outcomes**. Each item reports the **confidence** with which the text has been matched to an **intent** and the relative **intents**.

---

#### Algorithm 4 The natural language service main algorithm

---

```

1: token  $\leftarrow$  access_web_service()
2: while current_state_update do
3:   if speech.isRecorded then
4:     process_speech()
5:   end if
6: end while
```

---



---

#### Algorithm 5 *process\_speech*() function

---

```

1: response  $\leftarrow$  send_http_post_request()
2: data  $\leftarrow$  parse_data(response)
3: if data.outcomes.intent  $\neq$  NULL then
4:   do_something(intent)
5: end if
```

---

data structure returns a JSON packet like this:

```

{
  "msg_id" : "de7f215d-6aa8-4d48-8ce5-7fb8e7c56f3f",
  "_text" : "move forward ten meters",
  "outcomes" : [ {
    "_text" : "move forward ten meters",
    "confidence" : 0.453,
    "intent" : "go_ahead",
```

```

    "entities" : {
      "distance" : [ {
        "type" : "value",
        "value" : 10,
        "unit" : "metre"
      } ],
      "on_off" : [ {
        "value" : "on"
      } ]
    }
  } ]
}

```

### 5.2.3 Evaluation

GLOVR has been tested on VR demos designed ad hoc, running on a MacBookPro (Core i7 4578U, GT 750M graphics card, 16GB RAM). Oculus DK2 headset has been chosen as VR display but there are no restrictions for a future use with the upcoming devices. First game scenarios have been made with ThreeJS, a Javascript WebGL-based graphic library, that easily interfaces with Oculus and runs on all web browsers. GLOVR has demonstrated to work properly with either Chrome and Safari web browsers, even though the limitations in the VR frame rendering on WebGL were visible.

Then we focused on a single demo, *Floating Islands*, an action game where the player explores a world made by three floating islands. The game was fully developed in Unity, populated with characters and assets modeled with Blender. The scene has been reduced in terms of polygons to assure the best experience with Oculus, taking in account the limited performance of the mobile graphics card in use for the evaluation. In the game, the three islands have been populated by threatening animal enemies, from which the player has to defend himself, otherwise he/she can loose power. The player can also collect coins and can jump from an island to another one activating tricks hidden in the scene.

All the interactions in *Floating Islands* are covered by the GLOVR controller and can be casted in four groups:

- *Movement Controls*. The player can move in different directions, jump and rotate.
- *Action Controls*. The player hits the enemies with a punch, whose the strength is proportional to the force imposed by the hand.

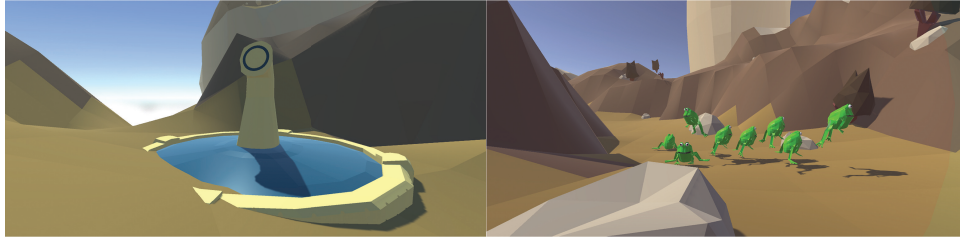


Figure 5.15: Two frames from *Floating Islands* demo.

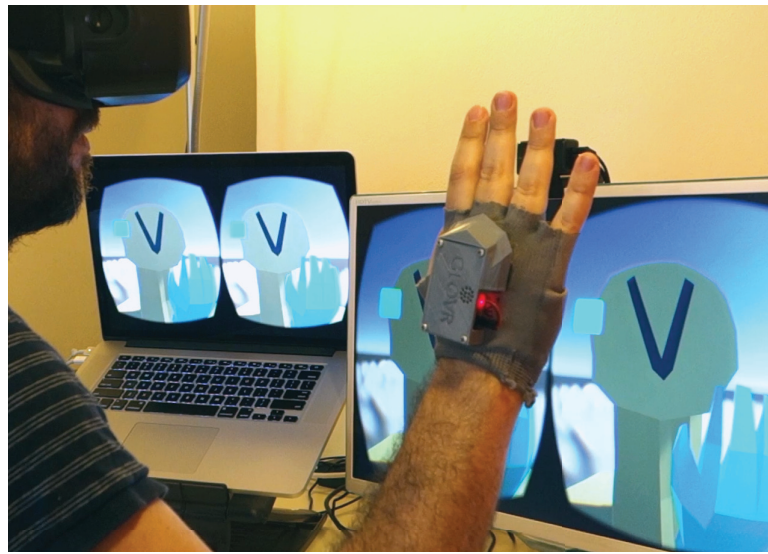


Figure 5.16: “V” gesture recognition.

- *Heads-Up Display (HUD) Controls.* Start/pause game, show/hide status, activate microphone.
- *Gesture Controls.* The player find in specific spots signals that suggest which gesture has to do in order to enable a connection between the islands.

We conducted our evaluation taking in account prior comprehensive research on 3D user interfaces, that featured similar characteristics [Kulshreshth and LaViola, 2015]. As for first comparison evaluation, we consider GLOVR Controls (GC) and a pair of Keyboard and Mouse Controls (KMC). At first glance KMC may appear an outdated configuration for VR environments. Actually it has been choose among the alternatives as the most efficient for emulating hand gestures, just typing selected keys and well known movements

with the mouse. Thus, in KMC the directions are taken with the directional keys and the character rotation with the mouse, as commonly implemented in classic game controls. Microphone activation and triggering of “O” and “V” gesture spots have been simulated with letter keys.

According to the design we proposed, we advanced the following hypotheses:

1. The user experience given by GLOVR is better than than playing with KMC configuration.
2. The motion sickness in movements due to the rotation of the character’s body in the virtual world is reduced.
3. Voice commands triggered by GLOVR controls won’t affect the game performance.
4. The gaming performance, measured in terms of killed enemies and life-time, is the same that we can aspect from a KMC configuration.

For the NL system we also measured the latency between the physical gesture of *Mic Activation* and the response given by Wit.ai, provided as a simple message on screen to prove the successful activation of the requested task. The participants were invited by a random external acoustic signal to ask two imperative sentences during the game session, related to some home automation task that ideally can be executed while playing a VR game:

- *Set the temperature in the bedroom to 24 degrees.* It simulates a command sent to an connected air conditioning.
- *Turn to day mode.* It changes the light in the game scene, switching from daylight to nightlight and vice versa.

We restricted the possible voice requests to a limited vocabulary. In fact, the NL classification accuracy of Wit.ai was not considered “under test”.

#### 5.2.4 Quantitative Method

For the evaluation of GLOVR we did a preliminary test with 12 participants (10 males, 2 females) in the age 20-38. At first they were briefly trained on the allowed control commands. All participants had prior experience with Virtual Reality thus we focused the usability test only with a VR version of Floating Islands, using with both GC and KMC configurations, and recording quantitative and qualitative measures with and without the use of the

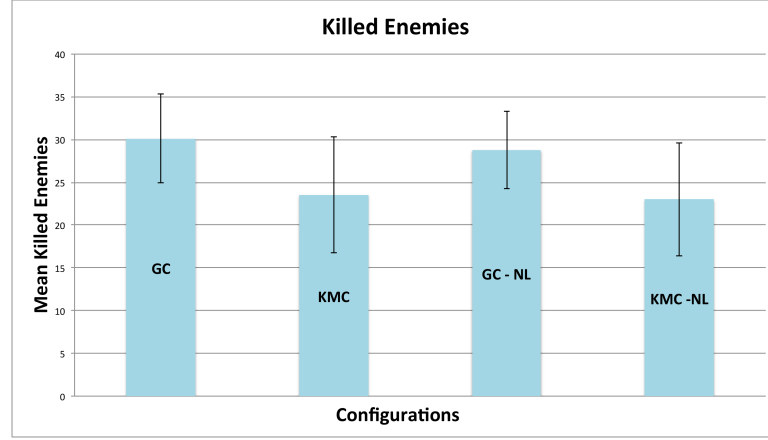


Figure 5.17: Mean number of killed enemies under different game conditions. GC: GLOVR Controls, KMC: Keyboard and Mouse Controls, GC-NL: GLOVR Controls with Natural Language activated, KMC-NL: Keyboard and Mouse Controls with Natural Language activated.

NL system. Because the participants played two trials for each mode, the total number of trials per participant was  $(GC + KMC) \times (NL + \text{not-NL}) \times (\text{Trial 1} + \text{Trial 2}) = 8$ . Each game had a maximum life time of 10 minutes. Performance was measured by means of a Two-Way ANOVA with *number of killed enemies* and *life time* as dependent variables. Testers killed more enemies ( $F_{1,44} = 13.42, p < 0.001$ ) with GLOVR (GC + not-NL configuration,  $\mu = 30.12, \sigma = 5.18$ ) than with common controls (KMC + not-NL,  $\mu = 23.54, \sigma = 6.78$ ). However, differences in number of killed enemies between NL and not-NL were not significant ( $F_{1,44} = 0.31, p = 0.58$ ). Also, no interaction between Controls and Natural Language System was found ( $F_{1,44} = 0.05, p = 0.81$ ). As expected, the life time was not affected by configurations ( $F_{1,44} = 0.11, p = 0.745$ ), much less by the presence of a NL system ( $F_{1,44} = 0, p = 0.9617$ ) (see Figure 5.17 and 5.18). The trials played with NL activated measured a mean response time of 2 seconds, that gives a total interval of 4 seconds from player's intent to triggering an action for GC (mean latency = 3.88,  $\sigma = 0.12$ ) and KMC (mean latency = 3.91,  $\sigma = 0.13$ ). As we expected the latency is dependent from the Wit.ai server, thus no significant relationship with the game configurations was observed (see Figure 5.19). Gesture recognition (GR) tasks were limited to the scene locations where the player could jump to another island. Two gestures were proposed, an "O" (or a circle) and a "V". Measures confirmed an expected accuracy around 90% ("O",  $\mu = 90.4, \sigma = 1.57$ , "V",  $\mu = 89.63, \sigma = 1.6$ )

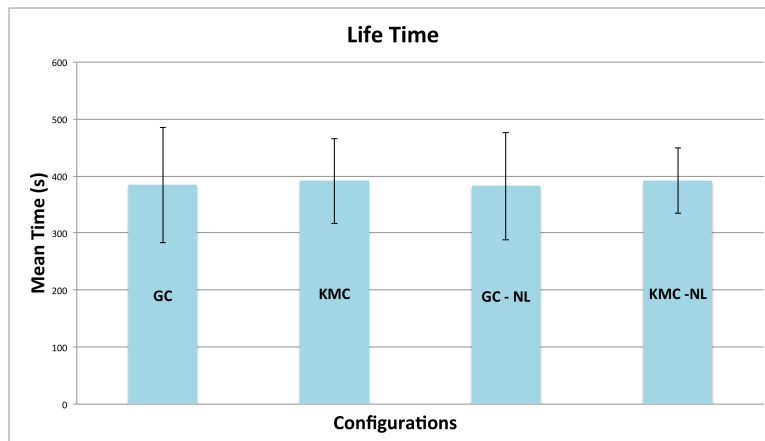


Figure 5.18: Mean time spent in a game session under the different conditions taken in account.

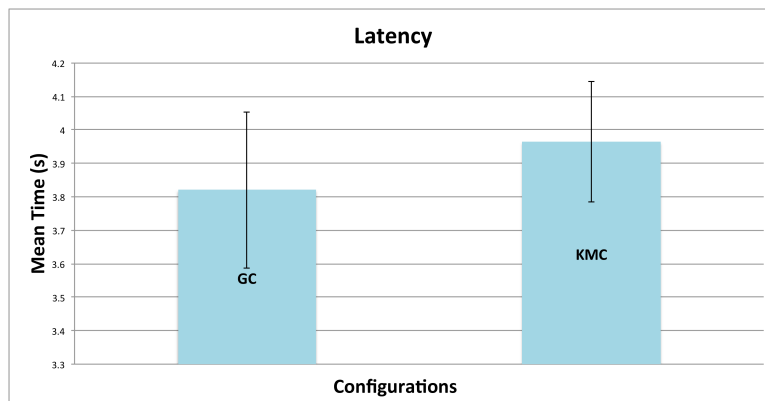


Figure 5.19: Mean Latency measured from the microphone activation to the response given by Wit.ai server.



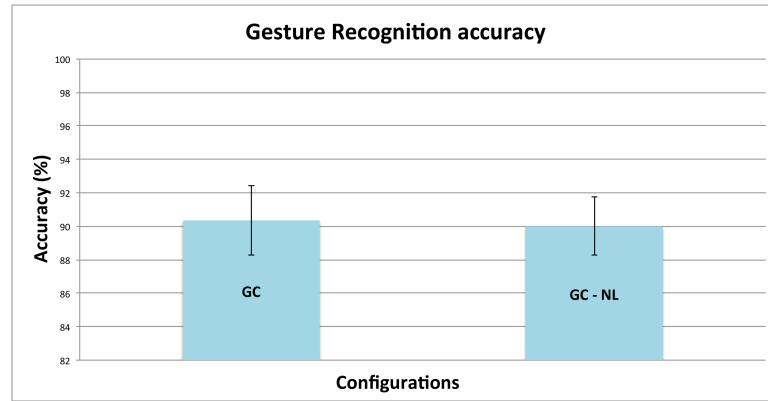


Figure 5.20: Classification accuracy measured for “O” and “V” gestures.

(see Figure 5.20).

### 5.2.5 Qualitative Method

After the trials, all the participants completed a 5 statements survey on a 5 point agreement Likert scale related to the overall game experience. We excluded specific questions for Gesture Recognition and Natural Language System features. The 5 questions were balanced to with an equal number of positive and negative statements to reduce the problem of acquiescence bias (see Table 5.1). The results of the survey confirmed the hypotheses: from Q1, users agreed that the game performance (killed enemies, time spent in a game session) was better, as reported by the measures. From Q2, we can say that GLOVR is comfortably wearable, despite the fact that device under test was just a “lab-made” prototype. A slightly positive opinion that GLOVR can reduce motion sickness came from Q3, given the more natural way to rotate the body of the character in the virtual world. Also, Q4 and Q5 feedbacks gave fairly positive opinions about the immersive experience and the ability of GLOVR to control directions properly (see Figure 5.21).

### 5.2.6 Discussion

We conducted a preliminary test that aimed to confirm hypotheses we previously advanced about GLOVR controller. The game performance we measured was globally the same as with traditional controllers and the users didn’t expressed particular concerns. However the game experience was

Survey Statements	
Q1	GLOVR performs better in terms of killed enemies and life time.
Q2	You feel comfortable for all the game session spent wearing GLOVR.
Q3	GLOVR controller seems reducing motion sickness.
Q4	GLOVR gives a less immersive experience.
Q5	GLOVR provides a difficult control of directions.

Table 5.1: The survey proposed to all the participants after the game session. The Likert scale goes from 1 (strongly disagree) to 5 (strongly agree).

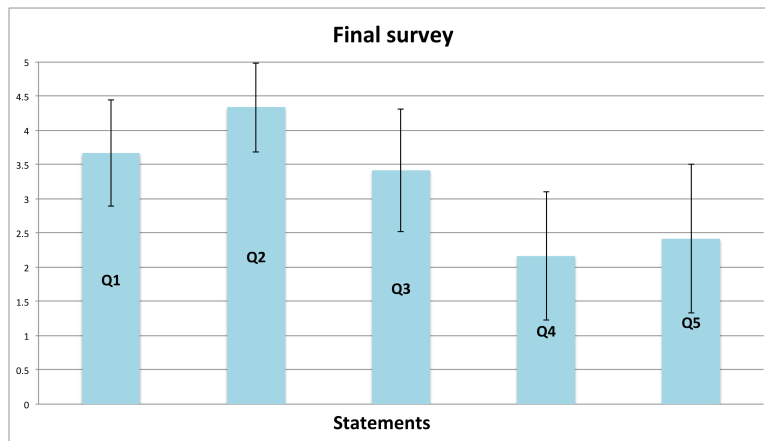


Figure 5.21: Results taken by the post-game survey.

judged more compelling and we guess that the main reason was the way users can play (jumping or punching, above all) more naturally. Survey results reported also a slight reduction of motion sickness, compared to KMC configuration. This result is just a first look in the *cybersickness* topic and of course needs further investigation. Exhaustive descriptions of the phenomena can be found in LaViola Jr. [LaViola, 2000] and Kolasinski [Kolasinski et al., 1995] works.

Future comparison tests will be made with game controllers specifically designed for VR applications as well as mixed configurations, that are theoretically possible, like GC + KMC or GC + GC (both hands).

Naturally, as in most VR applications, the connection between user's movements and consequent sickness represents an important aspect of our work. From this point of view, differently from other peripherals that are "calibrated" with the user's strength, GLOVR movements are strictly correlated to the force applied by the player, with the consequence that cybersickness is physiologically limited by the user him/herself while is playing.

The NL system implemented in GLOVR showed an average processing time of four seconds from the voice command to the response. This interval of time can be considered critical for taking immediate actions in games, but it is reasonably suitable for categories and external applications which not require instant feedback (strategy, casual games, simulations). From measures we cannot report any particular influence of the NL system in the overall game performance. We hypothesized that GLOVR shouldn't improve the game performance in terms of killed enemies because of the familiarity of GC + KMC compared to a wearable controller. On the contrary, some benefits have been observed. We can guess that the reduced motion sickness can be a significant factor for improving the quality of the game experience and also the overall game performance. However, any consideration about motion sickness need further and more appropriate tests.

### 5.2.7 Future Work

GLOVR, has been developed and successfully tested but, as the whole area of interest is evolving quickly, several enhancements can improve the user experience. GLOVR acquires data from an IMU and a microphone and allows the user to interact with the virtual world through gestures and natural language commands. This latter type of control allows interact with external applications during a VR game session, giving the user a more complete and satisfactory experience, without compromising his/her engagement. In a nutshell, we can summarize the key features and novelties of GLOVR as following:

1. GLOVR does not have any fingers and no fingers data are acquired. Despite the absence of their correspondence on the virtual environment, we opted for such a solution as a good compromise between wearability and richness of data. In any case a future version of GLOVR will feature also a tracking system in the way of Leap Motion and other trackers. The user can wear GLOVR for a long time with limited welding on the palm and without stress effects. Furthermore, he/she can handle almost any real object (game plugs?) while wearing GLOVR, allowing to import it in the virtual scene.
2. GLOVR acquires hand information by means of IMU sensors, thus it is not affected by occlusion issues present in systems like Leap Motion or Kinect or VR controllers such as Oculus Touch. Although the data are limited to the palm, past controllers like Wii Remote have showed how the palm controls can offer a compelling game experience and provide robustness in data. It is evident that all computer vision-based tracking suffer of occlusion issues and only a multicamera setup can guarantee continuous tracking. In any case a *IMU + Trackers* combo could be an interesting solution able to further enhance the user experience. It is evident that the absence of any positional information for GLOVR has to be compensated from other input devices. Particularly, Leap Motion or similar depth sensor can be integrated and offer a complete and more robust gesture recognition system. Further investigations will be done in this way.
3. All the actions in demos are easily playable and they don't need a particular training. The idea is to give to GLOVR a set of simple hand gestures that does not need any visual feedback for being reproduced correctly.
4. GLOVR features an easily gesture activable microphone. We have not seen so far significant examples of how to interact with external world (messaging, home automation, etc.) in VR. GLOVR and its natural language implementation represents an easy solution.

As for future work, the possibility to connect GLOVR with external home automation devices will be investigated. The exploitation of gesture interfaces to control smart objects has been explored with variable success [Fleer and Leichsenring, 2012] [Starner et al., 2000]. Natural Language solutions that allow faster response and a wider range of domains (like Api.ai) are under consideration. Furthermore, offline services have been taken in account. A first test was made with Pocket Sphinx, a offline version of CMU

Sphinx, that gave poor results because of the very limited vocabulary of the speech recognition system [Huggins-daines et al., 2006]. Although originally designed for Virtual Reality game applications, GLOVR can work in a variety of contexts, such as training, marketing, simulations, home automation, providing control outside VR as well. Future demos will be developed to cover significant opportunities in such fields.

# Chapter 6

## Conclusions

As long as the technology involved in Human-Computer Interaction has evolved, the mutual influence between Reality and Virtuality has become stronger. Data appeared like the “substance” that feed both the extremes of the Virtuality Continuum, making analog and digital worlds closer and sometimes indistinguishable. We focused on the design and implementation of technologies mostly related to entertainment and education, the areas where the immersive experiences are familiar to a larger audience. However, Mixed Reality found several promising applications in industrial and medical fields as well, and recent results confirmed this trend. In the course of our research we experimented some of the most recent Human-Computer Interfaces, as seen in details in Chapters 3,4 and 5.

### 6.1 Discussion

BRAVO started from the results achieved by previous research with consumer Brain-Computer Interfaces, we developed a system for the fruition of multimedia contents in a mobile device where the composition of text and images is based on the user’s brain activity. We designed different approaches, at first considering only the threshold reached in attention and meditation levels, then, according to a Computerized Adaptive Testing algorithm, we implemented a probabilistic solution. Evaluation with different groups of testers highlighted the potential for such technology in targeting more effectively the proposed learning content. We still consider consumer BCI “at early stage”, as the attention (and other cognitive processes) appear difficult to be decomposed. Sources of attention in public places are so numerous that it is really difficult to guess where the user pays attention.

Augmented Graphics (AG) origins from the need of a fast object recognition algorithm that could detect objects that are “similar” to the sample one. Computer vision recognition tasks are mostly based on a machine learning process that requires the acquisition of a dataset, the training of the recognition model according on it, and the classification of target images based on the highest generalization achieved by the model. We saw how AG focused on a very simplified model that consider only the distribution of features in the sample object, taking in account moments invariants theory for detecting such distribution on target objects. In these terms, AG accuracy cannot be compared with Convolutional Neural Networks or other learning approaches, because its goal is to group together any objects that “look like”. This approach is finding interest in gaming and educational contexts where the concept of “similarity” can be exploited and need to be light and portable. For this reason AG has been implemented in different demos on conferences and public events, showing solutions from interactive graphics novels to mobile games for kids.

In Chapter 5 we discussed GLOVR, a wearable hand controller that we judged a robust alternative to 3d tracking user interfaces, particularly suitable for Virtual Reality applications. It features an inertial unit that allows a continuous hand pose estimation and a the application of a gesture recognition system, a microphone for the implementation of a natural language system for sending voice commands to execute external tasks or interacting within the VR application.

Summarizing the main contributions, we saw with BRAVO how to exploit the user’s brain activity as an additional information for customizing contents, in a form that has not been explored before: an interactive system outside the typical laboratory context for EEG analysis.

We designed Augmented Graphics with some important assumptions in what we consider an “object”. Our approach, that cannot be considered a tracking solution, performs discretely in terms of detection rate but looks promising in what we expected it should do: recognize similar objects without any training/classification system.

For GLOVR we focused on the design of a wearable controller that could be useful not only for VR but also for other applications. Even if the glove concept is not true, we think that we targeted some of the most interesting aspects of VR controllers: high wearability, robustness on hand movement acquisition, thanks to the IMU sensors, ease to use because it doesn’t need any training, application in VR or not-VR contexts.

## 6.2 Future Work

We already anticipated our future directions in the previous chapters, but it is worth to spend a few other words in conclusion. Mixed Reality is “dimensionally” growing in terms of flux of information and it is more pervasive day by day thanks to the diffusion of wearable technologies that moved Augmented and Virtual Reality from the labs to the daily use. A world where we are connected in both real and virtual nature looks closer than we thought and the investments of technology industry in AR and VR show the strong interest in creating hybrid experiences, where people are immersed in scenarios where they can meet each other virtually and create real networks. In such context we are planning to develop new applications for the existing devices and to design interfaces that further exploit the mutual influence between humans and machines. We think that the popularity of Head Mounted Displays for VR and AR will make easy to implement widely other interfaces that so far have been limited to researchers or tech supporters. For instance, a Brain-Computer Interface could be easily integrated and useful for acquiring information about the user’s status and his/her cybersickness. We are working on the first version of a EEG board that can be inserted in the Razer’s OSVR headset. Depth cameras integrated in the HMDs can scan more and more precisely and objects taken from the reality can be “exported” to the virtual worlds. For this reason a natural progress for Augmented Graphics would be the extension to the third dimension. Controllers for the Mixed Reality have to be more natural, leaving behind traditional systems. To achieve this, devices need to be robust on data. We think that an integration of GLOVR with positional trackers could improve significantly the user experience. Furthermore, Natural Language services can provide tools for creating vocal personal assistants with minimum efforts. Such field of interest, so far a privilege of a few tech companies, will grow as long as Artificial Intelligence is expanding its horizon of applications. For GLOVR but also for other interfaces, we expect to make speech a smart controller itself.

## 6.3 Acknowledgements

I would like to thank: my PhD Advisor, Prof. Bruno Riccò, that has followed me through this adventure for three years. His experience in research and technology, the comments he gave me but also his humanity have been fundamental for doing better and better every day; my colleagues at Micrel Lab, for the long discussions about hardware and for the help on developing GLOVR; Antonella Guidazzoli from CINECA, for the contribution on



3D models used on BRAVO and for the suggestions she gave me across the PhD. Sofia Pescarin, from CNR-ITABC, for her incredible support of BRAVO and my other projects since ArcheoVirtual 2012; my reviewers, Prof. Mark Billingham from the University of South Australia, and Prof. Fotis Liarokapis from Masaryk University, for their precious help on giving clarity to this document. Last, but not least, all the professors, researchers, engineers and also the simple users that I met online and in all the conferences and public events I attended. The discussions with them and their feedback have enriched my vision and will contribute to my future work.

# Bibliography

- [Alcantarilla et al., 2013] Alcantarilla, P. F., Nuevo, J., and Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *In British Machine Vision Conference (BMVC)*.
- [Anderson et al., 2011] Anderson, J., Betts, S., Ferris, J., Fincham, J., and Yang, J. (2011). Using brain imaging to interpret student problem solving. *Intelligent Systems, IEEE*, 26(5):22–29.
- [API.ai, ] API.ai. Api.ai - build intelligent speech interfaces for apps, devices and web.
- [Baba and Asada, 2003] Baba, M. and Asada, N. (2003). Shadow removal from a real picture. In *ACM SIGGRAPH 2003 Sketches & Applications*, SIGGRAPH '03, pages 1–1, New York, NY, USA. ACM.
- [Beal et al., 2007] Beal, C., Mitra, S., and Cohen, P. R. (2007). Modeling learning patterns of students with a tutoring system using hidden markov models. In *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pages 238–245, Amsterdam, The Netherlands, The Netherlands. IOS Press.
- [Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522.
- [Berg and Malik, 2001] Berg, A. and Malik, J. (2001). Geometric blur for template matching. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–607–I–614 vol.1.
- [Berg et al., 2005] Berg, A. C., Berg, T. L., and Malik, J. (2005). Shape matching and object recognition using low distortion correspondences.

- In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 26–33, Washington, DC, USA. IEEE Computer Society.
- [Billingham et al., 2015] Billingham, M., Clark, A., and Lee, G. (2015). A survey of augmented reality. *Found. Trends Hum.-Comput. Interact.*, 8(2-3):73–272.
- [Bober, 2001] Bober, M. (2001). Mpeg-7 visual shape descriptors. *IEEE Trans. Cir. and Sys. for Video Technol.*, 11(6):716–719.
- [Bradski, 2000] Bradski, G. (2000). *Dr. Dobb's Journal of Software Tools*.
- [Calonder et al., 2010] Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: Binary robust independent elementary features. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Computer Vision ? ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 778–792. Springer Berlin Heidelberg.
- [Cam and Yang, 2000] Cam, L. and Yang, G. (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer New York.
- [Cameron et al., 2011] Cameron, C., DiValentin, L., Manaktala, R., McElhaney, A., Nostrand, C., Quinlan, O., Sharpe, L., Slagle, A., Wood, C., Zheng, Y. Y., and Gerling, G. (2011). Using electroactive polymers to simulate the sense of light touch and vibration in a virtual reality environment. In *Systems and Information Engineering Design Symposium (SIEDS), 2011 IEEE*, pages 121–126.
- [Carmona and Climent, 2012] Carmona, J. and Climent, J. (2012). A performance evaluation of hmm and dtw for gesture recognition. In Alvarez, L., Mejail, M., Gomez, L., and Jacobo, J., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 7441 of *Lecture Notes in Computer Science*, pages 236–243. Springer Berlin Heidelberg.
- [Castellani et al., 2014] Castellani, G., Remondini, D., and Intrator, N. (2014). Systems biology and brain activity in neuronal pathways by smart device and advanced signal processing. *Frontiers in Genetics*, 5(253).
- [Chan, 1996] Chan, T. (1996). Optimal output-sensitive convex hull algorithms in two and three dimensions. *Discrete and Computational Geometry*, 16(4):361–368.

- [Chen et al., 2004] Chen, Q., Petriu, E., and Yang, X. (2004). A comparative study of fourier descriptors and hu’s seven moment invariants for image recognition. In *Electrical and Computer Engineering, 2004. Canadian Conference on*, volume 1, pages 103–106 Vol.1.
- [Cheong et al., 2014] Cheong, H., Li, W., Shu, L., Bradner, E., and Iorio, F. (2014). Investigating the use of controlled natural language as problem definition input for computer-aided design. In *Innovative Design and Manufacturing (ICIDM), Proceedings of the 2014 International Conference on*, pages 65–70.
- [Cirett Galán and Beal, 2012] Cirett Galán, F. and Beal, C. (2012). Eeg estimates of engagement and cognitive workload predict math problem solving outcomes. In Masthoff, J., Mobasher, B., Desmarais, M., and Nkambou, R., editors, *User Modeling, Adaptation, and Personalization*, volume 7379 of *Lecture Notes in Computer Science*, pages 51–62. Springer Berlin Heidelberg.
- [Coulton et al., 2011] Coulton, P., Wylie, C. G., and Bamford, W. (2011). Brain interaction for mobile games. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek ’11, pages 37–44, New York, NY, USA. ACM.
- [Courgeon et al., 2014] Courgeon, M., Rautureau, G., Martin, J.-C., and Grynszpan, O. (2014). Joint attention simulation using eye-tracking and virtual humans. *Affective Computing, IEEE Transactions on*, 5(3):238–250.
- [Coursera, 2016] Coursera (2016). Free online courses from top universities.
- [Crowley et al., 2010] Crowley, K., Sliney, A., Pitt, I., and Murphy, D. (2010). Evaluating a brain-computer interface to categorise human emotional response. In *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*, pages 276–278.
- [Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
- [Deterding et al., 2011] Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining ”gamification”. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek ’11, pages 9–15, New York, NY, USA. ACM.

- [Ding et al., 2008] Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E. (2008). Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proc. VLDB Endow.*, 1(2):1542–1552.
- [Douglas and Peucker, 1973] Douglas, D. H. and Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122. doi:10.3138/FM57-6770-U75U-7727.
- [Dudani et al., 1977] Dudani, S. A., Breeding, K. J., and McGhee, R. (1977). Aircraft identification by moment invariants. *Computers, IEEE Transactions on*, C-26(1):39–46.
- [Fei-Fei et al., 2007] Fei-Fei, L., Fergus, R., and Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.*, 106(1):59–70.
- [Ferrari et al., 2006] Ferrari, V., Tuytelaars, T., and Van Gool, L. (2006). Simultaneous object recognition and segmentation by image exploration. In Ponce, J., Hebert, M., Schmid, C., and Zisserman, A., editors, *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 145–169. Springer Berlin Heidelberg.
- [Finlayson et al., 2006] Finlayson, G., Hordley, S., Lu, C., and Drew, M. (2006). On the removal of shadows from images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1):59–68.
- [Fleer and Leichsenring, 2012] Fleer, D. and Leichsenring, C. (2012). Miso: A context-sensitive multimodal interface for smart objects based on hand gestures and finger snaps. In *Adjunct Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST Adjunct Proceedings '12, pages 93–94, New York, NY, USA. ACM.
- [Flusser and Suk, 2006] Flusser, J. and Suk, T. (2006). Rotation moment invariants for recognition of symmetric objects. *Image Processing, IEEE Transactions on*, 15(12):3784–3790.
- [Folcher et al., 2014] Folcher, M., Oesterle, S., Zwicky, K., Thekkotttil, T., Heymoz, J., Hohmann, M., Christen, M., Daoud El-Baba, M., Buchmann, P., and Fussenegger, M. (2014). Mind-controlled transgene expression by a wireless-powered optogenetic designer cell implant. *Nat Commun*, 5.

- [G. Nikolakis, 2004] G. Nikolakis, D. Tzovaras, S. M. M. S. (2004). Cyber-grasp and phantom integration: Enhanced haptic access for visually impaired users. In *Conference Speech and Computer, St. Petersburg, Russia*, pages 507–513.
- [Geman et al., 1997] Geman, D., Amit, Y., and Wilder, K. (1997). Joint induction of shape features and tree classifiers.
- [Ghiani et al., 2015] Ghiani, G., Manca, M., and Paternò, F. (2015). Dynamic user interface adaptation driven by physiological parameters to support learning. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '15, pages 158–163, New York, NY, USA. ACM.
- [Gillian et al., 2011] Gillian, N., Knapp, B., and O'Modhrain, S. (2011). Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 337–342, Oslo, Norway.
- [Google, 2016] Google (2016). Google trends.
- [Graham, 1972] Graham, R. L. (1972). An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set. *Information Processing Letters*, 1:132–133.
- [Hadid et al., 2014] Hadid, A., Ylioinas, J., and Lopez, M. (2014). Face and texture analysis using local descriptors: A comparative analysis. In *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, pages 1–4.
- [Hammond, 2008] Hammond, B. (2008). A computer vision tangible user interface for mixed reality billiards. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 929–932.
- [Hammond and Davis, 2007] Hammond, T. and Davis, R. (2007). Ladder, a sketching language for user interface developers. In *ACM SIGGRAPH 2007 Courses*, SIGGRAPH '07, New York, NY, USA. ACM.
- [Hasan and Abdul-Kareem, 2014] Hasan, H. and Abdul-Kareem, S. (2014). Static hand gesture recognition using neural networks. *Artificial Intelligence Review*, 41(2):147–181.

- [Héloir et al., 2006] Héloir, A., Courty, N., Gibet, S., and Multon, F. (2006). Temporal alignment of communicative gesture sequences. *Computer Animation and Virtual Worlds (selected best papers from CASA '06)*, 17:347–357.
- [Hilbert, 1994] Hilbert, D. (1994). *Theory of Algebraic Invariants*. Cambridge University Press, New York, NY, USA.
- [Hillaire et al., 2012] Hillaire, S., Lecuyer, A., Regia-Corte, T., Cozot, R., Royan, J., and Breton, G. (2012). Design and application of real-time visual attention model for the exploration of 3d virtual environments. *Visualization and Computer Graphics, IEEE Transactions on*, 18(3):356–368.
- [Hu, 1962] Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187.
- [Huang and Leng, 2010] Huang, Z. and Leng, J. (2010). Analysis of hu’s moment invariants on image scaling and rotation. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 7, pages V7–476–V7–480.
- [Huggins-daines et al., 2006] Huggins-daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M., and Rudnick, A. I. (2006). Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *in Proceedings of ICASSP*.
- [Hughes et al., 2005] Hughes, C., Stapleton, C., Hughes, D., and Smith, E. (2005). Mixed reality in education, entertainment, and training. *Computer Graphics and Applications, IEEE*, 25(6):24–30.
- [Jarvis, 1973] Jarvis, R. (1973). On the identification of the convex hull of a finite set of points in the plane. *Information Processing Letters*, 2(1):18 – 21.
- [Jin, 2012] Jin, S. (2012). Design of an online learning platform with moodle. In *Computer Science Education (ICCSE), 2012 7th International Conference on*, pages 1710–1714.
- [Johns and Woolf, 2006] Johns, J. and Woolf, B. (2006). A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1, AAAI’06*, pages 163–168. AAAI Press.

- [Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45.
- [Kellaway, 1990] Kellaway, P. (1990). An orderly approach to visual analysis: Characteristics of the normal eeg of adults and children. *Current Practice of Clinical Electroencephalography*, pages 139–199.
- [Kim et al., 2001] Kim, Y., Kim, H., Ko, H., and Kim, H. (2001). Psychophysiological changes by navigation in a virtual reality. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 4, pages 3773–3776 vol.4.
- [Klein and Murray, 2009] Klein, G. and Murray, D. (2009). Parallel tracking and mapping on a camera phone. In *Proc. Eighth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’09)*, Orlando.
- [Kolasinski et al., 1995] Kolasinski, E., for the Behavioral, U. A. R. I., and Sciences, S. (1995). *Simulator sickness in virtual environments*. Number v. 4, no. 1027 in Technical report (U.S. Army Research Institute for the Behavioral and Social Sciences). U.S. Army Research Institute for the Behavioral and Social Sciences.
- [Koleva et al., 2000] Koleva, B., Schnädelbach, H., Benford, S., and Greenhalgh, C. (2000). Traversable interfaces between real and virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’00, pages 233–240, New York, NY, USA. ACM.
- [Kulshreshth and LaViola, 2015] Kulshreshth, A. and LaViola, Jr., J. J. (2015). Exploring 3d user interface technologies for improving the gaming experience. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, pages 125–134, New York, NY, USA. ACM.
- [Lamdan et al., 1990] Lamdan, Y., Schwartz, J. T., and Wolfson, H. (1990). Affine invariant model-based object recognition. *Robotics and Automation, IEEE Transactions on*, 6(5):578–589.
- [LaValle et al., 2014] LaValle, S., Yershova, A., Katsev, M., and Antonov, M. (2014). Head tracking for the oculus rift. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 187–194.



- [LaViola, 2000] LaViola, Jr., J. J. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56.
- [Lecuyer et al., 2008] Lecuyer, A., Lotte, F., Reilly, R., Leeb, R., Hirose, M., and Slater, M. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41(10):66–72.
- [Lee et al., 2002] Lee, J., Chai, J., Reitsma, P. S. A., Hodgins, J. K., and Pollard, N. S. (2002). Interactive control of avatars animated with human motion data. *ACM Trans. Graph.*, 21(3):491–500.
- [Lee et al., 2015] Lee, P.-W., Wang, H.-Y., Tung, Y.-C., Lin, J.-W., and Valstar, A. (2015). Transection: Hand-based interaction for playing a game within a virtual reality game. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 73–76, New York, NY, USA. ACM.
- [Lee et al., 2009] Lee, S., Jounghyun Kim, G., and Choi, S. (2009). Real-time tracking of visually attended objects in virtual environments and its application to lod. *Visualization and Computer Graphics, IEEE Transactions on*, 15(1):6–19.
- [Leeb et al., 2007] Leeb, R., Lee, F., Keinrath, C., Scherer, R., Bischof, H., and Pfurtscheller, G. (2007). Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 15(4):473–482.
- [Li et al., 2011] Li, Y., Li, X., Ratcliffe, M., Liu, L., Qi, Y., and Liu, Q. (2011). A real-time eeg-based bci system for attention recognition in ubiquitous environment. In *Proceedings of 2011 International Workshop on Ubiquitous Affective Awareness and Intelligent Interaction*, UAII '11, pages 33–40, New York, NY, USA. ACM.
- [Liarokapis et al., 2014] Liarokapis, F., Debattista, K., Vourvopoulos, A., Petridis, P., and Ene, A. (2014). Comparing interaction techniques for serious games through brain-computer interfaces: A user perception evaluation study. *Entertainment Computing*, 5(4):391 – 399.
- [Liarokapis et al., 2013] Liarokapis, F., Vourvopoulos, A., Ene, A., and Petridis, P. (2013). Assessing brain-computer interfaces for controlling serious games. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2013 5th International Conference on*, pages 1–4.

- [Lillo et al., 2010] Lillo, A. D., Motta, G., Thomas, K., and Storer, J. A. (2010). Shape recognition, with applications to a passive assistant. In Makedon, F., editor, *PETRA*, ACM International Conference Proceeding Series. ACM.
- [Ling and Jacobs, 2007] Ling, H. and Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):286–299.
- [Liu et al., 2013] Liu, N.-H., Chiang, C.-Y., and Chu, H.-C. (2013). Recognizing the degree of human attention using eeg signals from mobile sensors. *Sensors*, 13(8):10273.
- [Lord, 1980] Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum Associates.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110.
- [Madgwick, 2010] Madgwick, S. O. (2010). An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Report x-io and University of Bristol (UK)*.
- [Magenat et al., 2015] Maguenat, S., Ngo, D. T., Zund, F., Ryffel, M., Noris, G., Rothlin, G., Marra, A., Nitti, M., Fua, P., Gross, M., and Sumner, R. (2015). Live texturing of augmented reality characters from colored drawings. *Visualization and Computer Graphics, IEEE Transactions on*, 21(11):1201–1210.
- [Mann and Fung, 2001] Mann, S. and Fung, J. (2001). Videoorbits on eyetap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real scene. In *Proceedings of the Second IEEE International Symposium on Mixed Reality*, pages 48–55.
- [Marchesi, 2012] Marchesi, M. (2012). Neu: How brain activity can change an animated scene. In *ACM SIGGRAPH 2012 Posters*, SIGGRAPH ’12, pages 75:1–75:1, New York, NY, USA. ACM.
- [Marchesi et al., 2011] Marchesi, M., Farella, E., Riccò, B., and Guidazzoli, A. (2011). Mobie: A movie brain interactive editor. In *SIGGRAPH Asia 2011 Emerging Technologies*, SA ’11, pages 16:1–16:1, New York, NY, USA. ACM.

- [Marchesi and Riccò, 2013a] Marchesi, M. and Riccò, B. (2013a). Augmented graphics for interactive storytelling on a mobile device. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA '13, pages 59:1–59:1, New York, NY, USA. ACM.
- [Marchesi and Riccò, 2013b] Marchesi, M. and Riccò, B. (2013b). Bravo: A brain virtual operator for education exploiting brain-computer interfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 3091–3094, New York, NY, USA. ACM.
- [Marchesi and Riccò, 2016] Marchesi, M. and Riccò, B. (2016). Glovr: A wearable hand controller for virtual reality applications. In *Proceedings of the 2016 Virtual Reality International Conference*, VRIC '16.
- [Mele et al., 2006] Mele, K., Maver, J., and Sue, D. (2006). Image categorization using local probabilistic descriptors. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 336–340.
- [Mikolajczyk and Schmid, 2005] Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630.
- [Mikolajczyk et al., 2003] Mikolajczyk, K., Zisserman, A., and Schmid, C. (2003). Shape recognition with edge-based features. In *Proceedings of the British Machine Vision Conference*, volume 2, pages 779–788.
- [Milgram and Kishino, 1994] Milgram, P. and Kishino, F. (1994). A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS.
- [Mokhtarian et al., 1996] Mokhtarian, F., Abbasi, S., and Kittler, J. (1996). Efficient and robust retrieval by shape content through curvature scale space. pages 35–42.
- [Moodle, 2016] Moodle (2016). Open source learning platform.
- [Mostow et al., 2011] Mostow, J., Chang, K.-M., and Nelson, J. (2011). Toward exploiting eeg input in a reading tutor. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, AIED'11, pages 230–237, Berlin, Heidelberg. Springer-Verlag.
- [Nguyen et al., 2014] Nguyen, A. M., Yosinski, J., and Clune, J. (2014). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *CoRR*, abs/1412.1897.

- [Nigam et al., 2013] Nigam, S., Deb, K., and Khare, A. (2013). Moment invariants based object recognition for different pose and appearances in real scenes. In *Informatics, Electronics Vision (ICIEV), 2013 International Conference on*, pages 1–5.
- [Nijholt et al., 2008] Nijholt, A., Tan, D., Allison, B., del R. Milan, J., and Graimann, B. (2008). Brain-computer interfaces for hci and games. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3925–3928, New York, NY, USA. ACM.
- [Nishimura and Behrmann, 2009] Nishimura, Mayu, S. S. and Behrmann, M. (2009). Development of object recognition in humans. *F1000 Biology Reports*, pages 1–56.
- [Novick, 1966] Novick, M. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1):1–18.
- [Osada et al., 2002] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2002). Shape distributions. *ACM Trans. Graph.*, 21(4):807–832.
- [Parshall et al., 2012] Parshall, C., Spray, J., Kalohn, J., and Davey, T. (2012). *Practical Considerations in Computer-Based Testing*. Statistics for Social and Behavioral Sciences. Springer New York.
- [Paulson and Hammond, 2008] Paulson, B. and Hammond, T. (2008). Paleosketch: Accurate primitive sketch recognition and beautification. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 1–10, New York, NY, USA. ACM.
- [Pereira et al., 2011] Pereira, B. O., Expedito, C., De Faria, F. F., and Vivacqua, A. S. (2011). Designing a game controller for motor impaired players. In *Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, IHC+CLIHC '11, pages 267–271, Porto Alegre, Brazil, Brazil. Brazilian Computer Society.
- [Piton-Gonçalves and Aluísio, 2012] Piton-Gonçalves, J. and Aluísio, S. M. (2012). An architecture for multidimensional computer adaptive test with educational purposes. In *Proceedings of the 18th Brazilian Symposium on Multimedia and the Web*, WebMedia '12, pages 17–24, New York, NY, USA. ACM.
- [Quang Minh Khiem et al., 2010] Quang Minh Khiem, N., Ravindra, G., Carlier, A., and Ooi, W. T. (2010). Supporting zoomable video streams

- with dynamic region-of-interest cropping. In *Proceedings of the First Annual ACM SIGMM Conference on Multimedia Systems*, MMSys '10, pages 259–270, New York, NY, USA. ACM.
- [Rabiner, 1989] Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.
- [Ramer, 1972] Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1(3):244 – 256.
- [Rebolledo-Mendez et al., 2010] Rebolledo-Mendez, G., de Freitas, S., Rojano-Caceres, J., and Garcia-Gaona, A. (2010). An empirical examination of the relation between attention and motivation in computer-based education: a modeling approach.
- [Rebolledo-Mendez et al., 2009] Rebolledo-Mendez, G., Dunwell, I., Martínez-Mirón, E. A., Vargas-Cerdán, M. D., Freitas, S., Liarokapis, F., and García-Gaona, A. R. (2009). Assessing neurosky’s usability to detect attention levels in an assessment exercise. In *Proceedings of the 13th International Conference on Human-Computer Interaction. Part I: New Trends*, pages 149–158, Berlin, Heidelberg. Springer-Verlag.
- [Reckase, 1974] Reckase, M. (1974). An interactive computer program for tailored testing based on the one-parameter logistic model. *Behavior Research Methods and Instrumentation*, 6(2):208–212.
- [Rekimoto and Nagao, 1995] Rekimoto, J. and Nagao, K. (1995). The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology*, UIST '95, pages 29–36, New York, NY, USA. ACM.
- [Ruble et al., 2011] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571.
- [Segall, 1996] Segall, D. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2):331–354.
- [Sharp et al., 2015] Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., Freedman, D., Kohli, P., Krupka, E., Fitzgibbon, A., and Izadi, S. (2015). Accurate, robust, and flexible real-time hand tracking. In *Proceedings of the*

*33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 3633–3642, New York, NY, USA. ACM.

- [Silva et al., 2014] Silva, A. C., Mattioli, L. R., Paula, G. d., Cardoso, A., Lamounier, E. A., Lima, G. F. M. d., Prado, P. R. M. d., and Ferreira, J. N. (2014). A strategy to present 2d information within a virtual reality application. In *Proceedings of the 2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications, ISDEA '14*, pages 143–147, Washington, DC, USA. IEEE Computer Society.
- [Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video google: a text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477 vol.2.
- [Sklansky, 1982] Sklansky, J. (1982). Finding the convex hull of a simple polygon. *Pattern Recogn. Lett.*, 1(2):79–83.
- [Sluzek, 1994] Sluzek, A. (1994). Shape identification using new moment-based descriptors. In *TENCON '94. IEEE Region 10's Ninth Annual International Conference. Theme: Frontiers of Computer Technology. Proceedings of 1994*, pages 314–318 vol.1.
- [Starner et al., 2000] Starner, T., Auxier, J., Ashbrook, D., and Gandy, M. (2000). The gesture pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Wearable Computers, The Fourth International Symposium on*, pages 87–94.
- [Suk and Flusser, 2004] Suk, T. and Flusser, J. (2004). Projective moment invariants. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1364–1367.
- [Sun et al., 2012] Sun, Z., Wang, C., Zhang, L., and Zhang, L. (2012). Query-adaptive shape topic mining for hand-drawn sketch recognition. In *Proceedings of the 20th ACM International Conference on Multimedia, MM '12*, pages 519–528, New York, NY, USA. ACM.
- [Sutherland, 1968] Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, AFIPS '68 (Fall, part I)*, pages 757–764, New York, NY, USA. ACM.

- [Sutton, 2013] Sutton, J. (2013). Air painting with corel painter freestyle and the leap motion controller: A revolutionary new way to paint! In *ACM SIGGRAPH 2013 Studio Talks*, SIGGRAPH '13, pages 21:1–21:1, New York, NY, USA. ACM.
- [Suzuki and be, 1985] Suzuki, S. and be, K. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32 – 46.
- [Taranta II et al., 2015] Taranta II, E. M., Simons, T. K., Sukthankar, R., and Laviola Jr., J. J. (2015). Exploring the benefits of context in 3d gesture recognition for game-based virtual environments. *ACM Trans. Interact. Intell. Syst.*, 5(1):1:1–1:34.
- [Thomas, 2012] Thomas, B. H. (2012). A survey of visual, mixed, and augmented reality gaming. *Comput. Entertain.*, 10(1):3:1–3:33.
- [ToyTalk, ] ToyTalk. Toytalk - imagination + conversation.
- [Vasudevan et al., 2015] Vasudevan, V., Kafai, Y., and Yang, L. (2015). Make, wear, play: Remix designs of wearable controllers for scratch games by middle school youth. In *Proceedings of the 14th International Conference on Interaction Design and Children*, IDC '15, pages 339–342, New York, NY, USA. ACM.
- [Verplaetse, 1996] Verplaetse, C. (1996). Inertial proprioceptive devices: Self-motion-sensing toys and tools. *IBM Systems Journal*, 35(3.4):639–650.
- [Vlachos et al., 2003] Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., and Keogh, E. (2003). Indexing multi-dimensional time-series with support for multiple distance measures. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 216–225, New York, NY, USA. ACM.
- [Vuori et al., 2001] Vuori, V., Laaksonen, J., Oja, E., and Kangas, J. (2001). Experiments with adaptation strategies for a prototype-based recognition system for isolated handwritten characters. *International Journal on Document Analysis and Recognition*, 3(3):150–159.
- [Wahi et al., 2012] Wahi, A., Palamsamy, C., and Sundaramurthy, S. (2012). Rotated object recognition-based on hu moment invariants using artificial neural system. In *Information and Communication Technologies (WICT), 2012 World Congress on*, pages 45–49.

- [Wei and Wu, 2013] Wei, H. and Wu, L. (2013). A line-context based object recognition method. In *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on*, pages 250–255.
- [Weiss, 1985] Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53(6):774–789.
- [WEISS and KINGSBURY, 1984] WEISS, D. J. and KINGSBURY, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4):361–375.
- [Wit.ai, 2016] Wit.ai (2016). Wit.ai - natural language for the internet of things.
- [Wolpaw et al., 2002] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791.
- [Xing et al., 2009] Xing, K., Huang, J., Xu, Q., and Wang, Y. (2009). Design of a wearable rehabilitation robotic hand actuated by pneumatic artificial muscles. In *Asian Control Conference, 2009. ASCC 2009. 7th*, pages 740–744.
- [Xu et al., 2012] Xu, R., Zhou, S., and Li, W. (2012). Mems accelerometer based nonspecific-user hand gesture recognition. *Sensors Journal, IEEE*, 12(5):1166–1173.
- [Yoon et al., 2013] Yoon, H., wook Park, S., Lee, Y.-K., and Jang, J.-H. (2013). Emotion recognition of serious game players using a simple brain computer interface. In *ICT Convergence (ICTC), 2013 International Conference on*, pages 783–786.
- [Yuan and Hui, 2008] Yuan, R. and Hui, W. (2008). Object identification and recognition using multiple contours based moment invariants. In *Information Science and Engineering, 2008. ISISE '08. International Symposium on*, volume 1, pages 140–144.
- [Zhang et al., 2014] Zhang, H., Zheng, Q., and Zheng, G. (2014). A new shadow removal algorithm based on susan and cielab color space. In *Proceedings of International Conference on Internet Multimedia Computing and Service, ICIMCS '14*, pages 222:222–222:225, New York, NY, USA. ACM.



- [Zhu et al., 2014] Zhu, Z., Branzoi, V., Wolverson, M., Murray, G., Vitovitch, N., Yarnall, L., Acharya, G., Samarasekera, S., and Kumar, R. (2014). Ar-mentor: Augmented reality based mentoring system. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 17–22.
- [Zimmerman et al., 1987] Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S., and Harvill, Y. (1987). A hand gesture interface device. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, CHI '87, pages 189–192, New York, NY, USA. ACM.