Universidade do Minho
Escola de Engenharia

Carlos César Loureiro Silva

Using Predictive and Descriptive Cognitive Models for Evaluation of Interactive Computing Systems

abril de 2019

Using Predictive and Descriptive Cognitive Models for Evaluation of Interactive Computing Systems

Carlos César Loureiro Silva

UMinho|2019

Universidade do Minho
Escola de Engenharia

Carlos César Loureiro Silva

Using Predictive and Descriptive Cognitive Models for Evaluation of Interactive Computing Systems

Tese de Doutoramento em Informática

Trabalho efetuado sob a orientação do
Professor Doutor José Creissac Campos
e do
Professor Doutor Jorge Almeida Santos

abril de 2019

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Licença concedida aos utilizadores deste trabalho

Agradecimentos

A escrita de uma Tese de Doutoramento é um processo curioso. Quase sempre penoso, frequentemente frustrante, mas que perto do fim se reveste de um prazer difícil de explicar. Celebrar a ultrapassagem dos muitos momentos difíceis deste processo, só faz sentido se puder ser partilhado com as pessoas da nossa vida.

Assim, sem que a ordenação seja sinónimo de importância, quero agradecer aos meus colegas e amigos de laboratório pelo apoio diário durante os últimos anos. Foram eles que viveram de perto este processo, vindo muitas vezes deles o apoio e a crítica construtiva que se revela fundamental. Ao Bruno, à Liliana, e à Sandra agradecer tudo aquilo que me ensinaram e o espetacular acolhimento de um miúdo lhes foi parar às mãos à uns 10 nos atrás. Ao Emanuel, aos Joões (Pedro e Lamas), e à Joana companheiros de luta dos últimos anos, a paciência, a ajuda, e o incrível profissionalismo de um grupo de jovens investigadores que foram capazes de criar um espaço de trabalho ao qual me orgulho de pertencer. To Paolo, Glenn, Yi Zhang, and Paul Jones I want to thank you the opportunity you gave me to work in a wonderful place and the good moments we enjoyed in Maryland.

Aos colegas do Centro de Computação Gráfica, instituição que me acolheu no início desta tese como bolseiro e que me permitiu terminá-la já como membro integrado da sua equipa de investigação.

Ao Pierre, ao Tiago, ao Diogo, ao Pedro e a Nocas, a amizade verdadeira que foi sempre reafirmada a cada oportunidade que me deram para escapar do trabalho e entrar num mundo sempre muito cómico e de uma loucura saudável. Uma das melhores coisas de terminar uma tese, é poder deixar de ter desculpas para recusar os convites dos amigos.

Aos meus orientadores, Professor José Creissac Campos e Professor Jorge Santos, por terem sempre acreditado em mim, ignorando sucessivos adiamentos para rematar isto e estando sempre presentes quando a necessidade de orientação surgia. Não podia ter pedido melhor orientação. Se um dia estiver no vosso papel, certamente imitarei a vossa postura esperando que compreendam que a imitação é a melhor forma de elogio.

À minha família, que mesmo não sabendo muito bem o que faço acreditam que é a coisa mais importante do mundo. Não é essa a melhor demonstração de carinho que alguém pode receber? Em tudo o que faço, o vosso apoio é fundamental.

Por último, à pessoa a quem dedico todo este trabalho esperando que a sua conclusão me permita mais facilmente retribuir todo o amor demonstrado ao longo destes últimos seis anos. À Mariana, o meu mais profundo obrigado e a promessa de recuperar para nós o tempo que mereces.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, 29/04/2019

Full Name: Carlos César Loureiro Silva

Signature:

Using Predictive and Descriptive Cognitive Models for Evaluation of Interactive Computing Systems

The advent of Interactive Computing Systems (ICSs) has released the great potential that computers have for influencing everyday human activity. The establishment of Human Computer Interaction (HCI) as an academic discipline has helped develop a sense that serious consideration and understanding of how humans interact with computers is a pressing necessity for both the computer scientist and the developer.

One of the traditional ways of using HCI research to foster the development of better technology is by resorting to the use of *Interaction Models* – or simplifications of the reality capable of being operationalized in such a way that provide valuable information in order to adapt a given system to its user. Two types of models, often respectively referred as *Predictive Cognitive Models* (PCMs) and *Descriptive Cognitive Models* (DCMs), have emerged from the HCI research in the last decades, and they have been used with the primary goal of formalizing empirical insights and foreseeing design and interaction problems.

Throughout this thesis, we will review the prolific history of user-research in HCI while showing contributions of both interaction-modeling approaches for the development of innovative and safer user interface technology. As a way of proving the benefit these approaches can bring to the development of today's ICSs, we will demonstrate their application to the evaluation and development of two types of ICSs that are at the forefront of current HCI challenges: *Immersive Virtual Environments* (IVEs), and *Safety-critical Interactive Medical Devices*. Although distinct in principle, these ICSs have some communalities that render them interesting use-cases in the scope of this thesis. Both systems are becoming increasingly widespread, and due to their particularities, both present new interaction challenges that are absent from more traditional HCI use-cases, such as the personal computer.

The main outputs of this thesis are thus twofold: the outline of new PCMs and design guidelines that can help the development of increasingly immersive virtual environments; and the outline of a new DCM of use-error with safety-critical devices in addition to a set of design guidelines that can help mitigate these errors and foster the development of increasingly safer medical devices.

Keywords: Descriptive Cognitive Models; Immersive Virtual Environments; Interactive Computing Systems; Predictive Cognitive Models; Safety-Critical Medical Devices.

Utilização de Modelos Cognitivos Preditivos e Descritivos para Avaliação de Sistemas Interativos de Computação

O surgimento de Sistemas Interativos de Computação (SIC) potenciou o impacto e a influência que os computadores exercem sobre a atividade quotidiana dos seres humanos. A consolidação da Interação Humano-Computador (IHC) como uma disciplina académica, ajudou a estabelecer a ideia de que a séria consideração e o estudo aprofundado da forma como os utilizadores interagem com os computadores, é uma necessidade premente tanto para o cientista da computação, como para o técnico que desenha e desenvolve novos SIC.

Uma das formas tradicionais de utilizar os métodos de IHC para fomentar o desenvolvimento de melhor tecnologia, é recorrendo ao uso de *Modelos de Interação* – simplificações da realidade capaz de serem operacionalizadas de modo a fornecer informações valiosas para adaptar um sistema ao seu utilizador. Da investigação em IHC emergiram dois tipos de modelos, frequentemente referidos como *Modelos Cognitivos Preditivos* (MCP) e *Modelos Cognitivos Descritivos* (MCD), usados com o objetivo principal de formalizar resultados empíricos e prever problemas de *design* da interação.

Ao longo desta tese, iremos rever a prolifica história da investigação com utilizadores em IHC e demonstrar como ambas as abordagens de modelação da interação apoiaram o desenvolvimento de tecnologia, segura e inovadora, de interface com o utilizador. De forma a demonstrar que estas abordagens podem trazer o mesmo benefício ao desenvolvimento dos SIC atuais, descreveremos a sua aplicação à avaliação e desenvolvimento de dois tipos de SIC, na vanguarda dos atuais desafios da IHC: *Ambientes Imersivos e Virtuais* (AIVs), e *Dispositivos Médicos Interativos e de Segurança-Crítica*. Embora distintos na sua gênese, estes SIC têm algumas semelhanças que os tornam casos de uso interessantes no âmbito desta tese. Ambos os sistemas estão cada vez mais difundidos e, devido às suas particularidades, apresentam novos desafios de interação que estão ausentes nos casos de uso tradicionais da IHC, como o computador pessoal.

Desta forma, os resultados desta tese dividem-se em duas principais áreas: o desenvolvimento de novos MCPs e diretrizes de *design* para apoiar o desenvolvimento de ambientes virtuais cada vez mais imersivos; e o desenvolvimento de um novo MCD de erro de uso em dispositivos de segurança-crítica, além de um conjunto de diretrizes de *design* que podem ajudar a mitigar esses erros e promover o desenvolvimento de dispositivos médicos cada vez mais seguros.

Palavras-Chave: Ambientes Imersivos Virtuais; Dispositivos Médicos Interativos e de Segurança-Crítica; Modelos Cognitivos Descritivos; Modelos Cognitivos Preditivos; Sistemas Interativos de Computação.

Table of Contents

PART I General Introduction

PART II PCMs and Immersive Virtual Environments

PART III DCMs and Safety-Critical Interactive Medical Devices

PART IV General Conclusions and the Future of ICSs

List of Figures

List of Tables

Abbreviations

Part I

UI	User Interfaces	PCM	Predictive Cognitive Model
SRI	Standford Research Institute	DCM	Descriptive Cognitive Models
NLS	oN-Line System	GUI	Graphical User Interfaces
ARC	Augmentation Research Center	EICS	Engineering Interactive Computing Systems
ACM	Association for Computing Machinery	ICS	Interactive Computing Systems
ICS	Interactive Computing System	IVE	Immersive Virtual Environment
HCI	Human-Computer Interaction	SBME	Simulation-Based Medical Education

Part II

HMD	Head Mounted Display	CRT	Cathode Ray Tube
VR	Virtual Reality	CRT	Choice Reaction Time
AR	Augmented Reality	KLM	Keystroke-Level Model
WTI	Window of Temporal Integration	MHP	Model Human Processor
RV	Reality-Virtuality	EPIC	Executive Process-Interactive Control
EWK	Extent of World Knowledge	CPS	Central Production System
AV	Augmented Virtuality	VIVED	Virtual Visual Environment Display
ID	Index of Difficulty	CGI	Computer-Generated Images

OLED	Organic LED display
FoV	Field of View
CAVE	Cave Automatic Virtual Environment
TORE	The Open Reality Experience
SPL	Sound Pressure Level
ITD	Interaural Time Difference
ILD	Interaural Level Difference
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
ISM	Image Source Method
RIR	Room Impulse Responses
FDA	Food and Drugs Administration
MAUDE	Manufacturer and User Facility Device Experience database
AECL	Atomic Energy of Canada Limited
ETCC	East Texas Cancer Center
CWA	Cognitive Work Analysis
GEMS	Generic Error-Modelling System
GOMS	Goals, Operators, Methods, Selection

WTI	Window of Temporal Integration
TOJ	Temporal Order Judgment
SOA	Stimulus Onset Asynchrony
PSS	Point of Subjective Simultaneity
PLW	Point Light Walker
VE	Ventriloquism Effect
POI	Point of Interest
JND	Just Noticeable Difference
UA	Unity Assumption
FMDT	Frontal Matching Distance Task
RWS	Real-World Scenario
IVEPH	Photorealistic Immersive Virtual Environment
IVENPH	Non-photorealistic Immersive Virtual Environment

Part III

CPS	Cyber-Physical Systems
ICE	Integrated Clinical Environment
HFCF	Human Factors Classification Framework
CPOE	Computerized Physician Order Entry
GEMS	Generic Error Modeling System
SBME	Simulation-Based Medical Education

Part I

General Introduction

1. THE REVOLUTION OF INTERACTION

There will always be plenty of things to compute in the detailed affairs of millions of people doing complicated things.

Vannevar Bush, *As We may Think*

1.1. At the Dawn of Computation

The term *Computeture* can be traced to the writings of the 1th century roman author Pliny the Elder when, in his *Naturalis Historia,* he describes how the breadth of Asia could be properly *calculated* – "…*latitudo sane computetur*" – from the Ethiopic Sea to Alexandria on the Nile [1]. By the first half of the 20th century, the word *computer* had already been in intensive use, primarily to mean a human activity (in the 18th and 19th century), later to name a profession (in the early 20th century), and in the 40's of the past century it started to be applied as a term to describe a new technology capable of performing calculations of a complexity and speed unachievable by its human counterpart. In the 50's, computers where still conceptualized as fundamentally powerful calculators and as such user interfaces (UIs) where merely a way of conveying instructions towards computing a given result. The first computers UI were essentially a programing interface and, frequently, a poorly usable one (see [2] for a concise history on electronic analog computing).

Nevertheless, in 1945 some computer scientists were already grasping the potential computers could have in influencing everyday human activity. In that year, Vannevar Bush, first dean of engineering at the MIT and one of the developers of the early analog computers, published a detailed description of the *Memex*[1] - a system for storing and retrieving scientific information. His description of this conceptual system placed a great focus on new mediums of user interaction such as screens, photography cameras to register and catalog new information, and speech-recognition systems for digital stenography (**Figure 1**).

[1] The description of the Memex system appeared first in the Atlantic Monthly on July 1945. In the previous years, Vannevar Bush had been the science advisor for President F.D. Roosevelt, playing that role during World War II. After the war, the great investment in military research would have to be redirect to new endeavors and, in Bush's opinion, a project to develop the Memex systems could be the ideal replacement for the Manhattan project, again joining the best minds of a generation only now for a peaceful purpose during a time of peace [4].

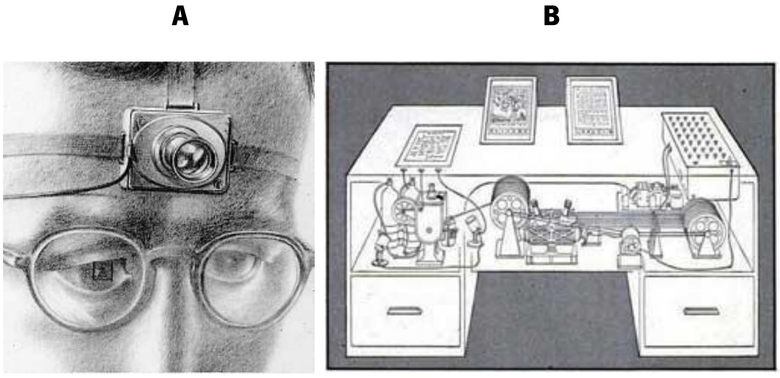


Figure 1. *Panel A*, rendition of the camera used by the scientist - Memex user – to capture and catalog new information. *Panel B*, a diagram of the Memex desktop-like system as described by Bush in his article entitled "As We May Think" [3]. Images from LIFE.

Although Bush's bold vision was certainly difficult to materialize at that time, it contributed to the idea of the computer as an enabler of information access, storing, and sharing amongst its users. Following Bush's vision[2], during the 1950's, Douglas Engelbart started a major research project at the Stanford Research Institute that included the development of new user interfaces for computer systems, of which the computer mouse was one of the components. The original concept was a unified system, which for the first time had in consideration all the interaction process ranging from displaying information to the user through a digital medium, to taking into account users' input and act upon that information (**Figure 2**). The final goal of the system, which became known as the *oN-Line System* (NLS), was to enable a wide variety of user interaction with digital information that, in Engelbart's words, would contribute to *"augmenting human intellect"*. Suitably, Englerbart's laboratory was named the *Augmentation Research Center* (ARC) and its vision was that of increasing the effectiveness of individuals through *the use* of computer digital systems in an easy and effective way. With a good measure of correct forecast, Engelbart argued that these systems could fundamentally change the way people collaborate, work, and ultimately think.

[2] Engelbart's reports of the NLS quoted extensively Bush's article, "As We May Think".

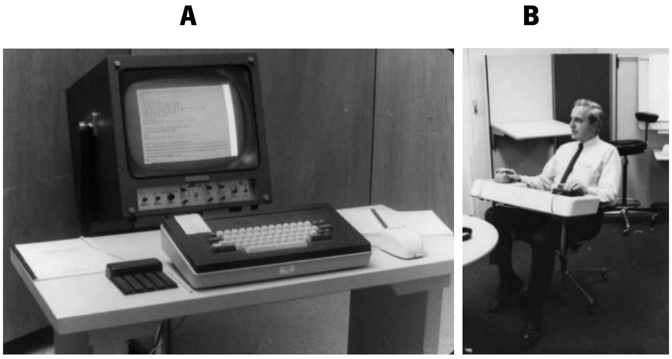


Figure 2. *Panel A,* shows one of the several developed desktop configurations for the NLS. All of these configurations included a Cathode-Ray Tube (CRT) display, a computer mouse, its complimentary chord set, and a keyboard. It is important to note how this arrangement is similar to today's desktop computers. *Panel B*, shows what had become the most popular mock-up design for the NLS (in this image operated by Engelbart), the Hermann Miller design. Images from Stanford.edu.

The NLS was presented in 1968 at the *Association for Computing Machinery* (ACM) *Fall Joint Computer Conference* in San Francisco, and Engelbart's demonstration of the overall system became known for posterity as "the mother of all demos"[3]. In addition to new user interfaces such as the computer mouse and its associated keyset, Engelbart's also showcased some early demonstrations of today's well-known technology such as video-conferencing, collaborative text editing, outlining tools, and hyperlinks [4]. In a way, Engelbart was the first to materialize successfully what we now know as an *Interactive Computing System* (ICS). Once serious consideration was taken on how these systems would be helping humans in everyday tasks, capturing and understanding how humans interact with computers became a pressing necessity for both computer scientists and developers.

1.2. The Contribution of Human-Computer Interaction

In its official curricula, the ACM defines *Human-Computer Interaction* (HCI) as:

> "… *a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.*" (p.5) [5].

As we will see some of the seminal works in HCI have emerged as early as in the mid 50's (see **Section 2.5** and the works of Paul Fitts), nonetheless HCI as a research field became formalized in the early 1980s, initially as a specialty area in computer sciences but rapidly expanded as an interdisciplinary

[3] The full version of the Demo can be viewed here: https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/1968Demo.html

field that incorporated diverse concepts and approaches from a collection of semi-distinct fields of research and practice in human-centered informatics. It is a matter of ongoing discussion if we should define HCI as an autonomous scientific discipline since it appears that some of its characteristics fail at *popperian* aspects of a scientific discipline definition, as it is the case of its multidisciplinary character, the lack of a clear methodology, and the weak separation between practice and research (see [6], [7] or [8] for a debate on this issue). Nevertheless, it is of common agreement that HCI provides, in the words of Yvonne Rogers [9], "*prescriptive knowledge*", giving advice on how to design an ICS. In addition, as pointed out by Alan Dix, there is an increasing sense of community within practitioners and researchers in computer and cognitive sciences that pursue two main common goals [10]:

1- Improve interactions between computers and their users by making them more useable and receptive to the users' needs (*primary goal*);

2- Design systems that reduce the barrier between the users' cognitive model of what they want to accomplish and the computer's understanding of the users' task (*long-term goal*).

In this sense, we can classify HCI as an academic discipline, despite knowing that HCI implies sometimes not just scientific, but also engineering and even craft work. Interestingly, the aspects that render HCI a peculiar epistemological case are also the ones that give HCI its strength and its potential to continually find innovative computational solutions. If we take, for instance, the unclear separation between craft, engineering and science and the frequent iterative nature of the work in HCI, we surely can view these characteristics as making part of a rather empirical approach as opposed to the more scientific traditional ones. At the same time, however, it is not difficult to acknowledge the advantages of such approach when the goal is to come out with an exemplary product, in terms of its usability. Similarly, if we consider the multidisciplinary character of HCI, we can always argue that it is responsible for broadening too much its academic frontiers and give HCI that typical aspect of an embryonic scientific area – the difficulty in defining itself. Nevertheless, we should also acknowledge that the coexistence between such different areas within the HCI umbrella has been one of its main practical advantages, since it allows the interconnection of previously unrelated knowledge and contributes for the advance of each one of these disciplines (forming the HCI universe) individually. According to the ACM curricula on human-computer interaction [5], different disciplines, such as cognitive psychology, sociology, human factors, product design, and industrial engineering, function as supporting disciplines for computer science in the scope of HCI, much as physics serves as a supporting discipline for civil engineering, or as mechanical engineering serve as a supporting discipline for robotics.

The design of many modern computer applications inevitably requires the design of components of the system that allow for interaction between the system and its user, and these components typically represents more than half a system's lines of code [5]. In the scope of computer science, HCI is responsible to guide us during the definition of the functionality a system will have, on how to bring it out to the user, and finally, on how to build and to test its design. One of the traditional ways of using HCI research to foster the development of better technology is by resorting to the use of *Interaction Models*. A model is a simplification of the reality capable of being operationalized in such a way that provides valuable information in order to adapt a given system to the subject being modeled. Their typology can be quite diverse and range from a formal language to a verbal-analytical description of behavior [11]. These two types of models are often respectively referred as *predictive* and *descriptive* models. HCI often uses both types of models with the goal of formalizing empirical insights and foreseeing design and interaction problems.

Some HCI research methodologies were specifically developed to evaluate user's interaction performance. The outputs of these research methodologies are often condensed in a *Predictive Cognitive Model* (PCM) based on the quantitative analysis of the user's behavior. A PCM is a model that predicts how a given piece of information (input) will affect human behavior (output), and might be represented through an equation that predicts the outcome of a dependent variable based on the value of one or more independent variables – or predictors [11]. In the historical context of the HCI discipline, the independent variable has often been a given displayed piece of information or a specific design choice for an input device, while the dependent variable is often the time or speed when performing a task, its accuracy, and/or error rate. Nevertheless, any dimension of human behavior is capable of serving as the output of a PCM, as long as it is feasible to be measured precisely on a continuous or ratio scale [11]. The final goal of PCMs is the prediction of human performance, but in order to achieve their goal these models should give account of how information is processed and what is the role of human perception, memory, attention, and motor skill (of the user) in the outcome.

Descriptive Cognitive Models (DCMs), on the other side, are applied more often during the conception or early development stages of a new interface, or during the study of use-error, as a "forensic tool". As MacKenzie puts it, *"they emerge from a process so natural it barely seems like modeling"* (p. 233) [11]. The process Mackenzie is referring to is the partition, analysis, and description of the interaction process. This purely analytical process has since long been conducted by experts, analysts, and developers in order to obtain insight on how the interaction flow, the design features, or the information content of an interface might lead to performance deficits, faulty interactions or use errors [12].

As we will see in **Chapter 2** and **Chapter 4**, history is prolific in showing us contributions of both HCI interaction-modeling approaches for the development of innovative and safer user interface technology. User studies using both quantitative and qualitative methodologies have helped spark the development of highly diversified interactive computing systems. As pointed out by John Carroll, until the late 1970s, the interaction with computers was reserved to information technology professionals and a small group of dedicated hobbyists. In this way, the need for HCI was kept hidden under the existence of expert users only [7]. Nevertheless, this was dramatically changed around 1980, when the emergence of personal computers – and the advent of direct manipulation on *graphical user interfaces* (GUI) – turned HCI into an indispensable tool, once the use of computer systems as consumer goods became widespread (thus, fulfilling the realization of Engelbart's vision)[4]. Today's interactive computer systems encompass a great variety of platforms ranging from the single-user desktop computer to the use of ubiquitous systems in public spaces. This is evident in the ACM conference on *Engineering Interactive Computing Systems* (EICS), which covers topics of interest so diverse as: *design and development of systems incorporating new interaction techniques and multimodal interaction; multi-device interaction*; *mobile and pervasive systems*; *safety-critical interactive devices; and immersive virtual environments* (among others).

The last two ICSs of the list above are at the forefront of today's HCI challenges[5] and will be the main use cases for the work developed in the scope of this thesis. Throughout the next chapters, we will be illustrating the contribution HCI can give to the evaluation and development of these two types of interactive computing systems:

1 *Immersive Virtual Environments* (IVEs), can be regarded as the epitome of interactive computer systems due to their ability to appeal to the different senses of the user during the interaction process. Nevertheless, knowledge about perception and performance – two fundamental steps

[4] Around the time of Engelbart's NLS, Ivan Sutherland was also providing a demonstration of direct manipulation interfaces in his 1963 PhD thesis on the Sketchpad, a system that supported the users' manipulation of graphic objects using a light-pen that could interact with the image displayed on a CRT. These seminal works and the advent of the firsts CAD/CAM systems (the General Motors one, in 1963) led to an all new range of technology and interface ideas, and soon the first commercial systems to make extensive use of direct manipulation interfaces were on the market: the Xerox Star in 1981, the PERQ in 1981, the Apple Lisa in 1982, and the Apple Macintosh in 1984.

[5] In addition to the EICS topics of interest, the top HCI conference, *Computer Human Interaction* (CHI), has a track dedicated to *Interaction techniques, Devices and Modalities* with mostly IVEs related topics such as: haptic and tangible interfaces, 3D interaction, augmented/mixed/virtual reality, wearable and on-body computing, sensors and sensing, displays and actuators, muscle- and brain-computer interfaces, and auditory and speech interfaces. Another important conference in HCI, INTERACT, has a special track on *safety/critical interactive systems* with topics as automation, critical interactive systems, healthcare, human error, human work interaction design, safety, security, space, training, transportation.

of the interaction process – in IVEs is still scarce and in need of standardized evaluation methodologies. Throughout this thesis, we will show how HCI empirical approaches can help us to develop new (or improve upon existing) *Predictive Cognitive Models (PCMs)* capable of providing useful design guidelines for developers looking to produce increasingly immersive IVEs.

2 *Safety-critical interactive devices,* such as information and control systems in power plants, cyber-physical medical systems, or medical devices, are of great importance for HCI research because errors during the interaction process (also referred to as use-errors) can have dramatic consequences for the user's safety. Increasingly detailed descriptions of the interaction process are, therefore, highly important to be able to predict and mitigate possible future use-errors. Throughout this thesis, a particular type of these systems – medical devices – will be our use case for presenting the benefit of using *Descriptive Cognitive Models (DCMs)* capable of functioning as a frame of reference for understanding usage behavior and providing useful design guidelines for developers to produce increasingly safer interfaces.

Although distinct in principle, IVEs and safety-critical interactive devices have some communalities that render them interesting use-cases in the scope of this thesis. Both systems are becoming increasingly widespread, and due to their particularities, both present new interaction challenges that are absent from more traditional HCI use-cases, such as the personal computer.

1.3. Approach

The main goal of this thesis is to demonstrate that the use of interaction-modeling methodologies can be applied to:

1. Gather knowledge about the user and its interaction process to construct models that are capable of describing/explaining/or predicting user behavior;

2. Develop standardized methodologies that support testing and validation of interactive computing systems.

IVEs and safety-critical systems are especially suitable use cases for applying different interaction-modeling methodologies. Scott Mackenzie, presents a *Continuum of Models* (**Figure 3**) that places interaction-modeling outputs along a continuum going from descriptive to predictive models [11].

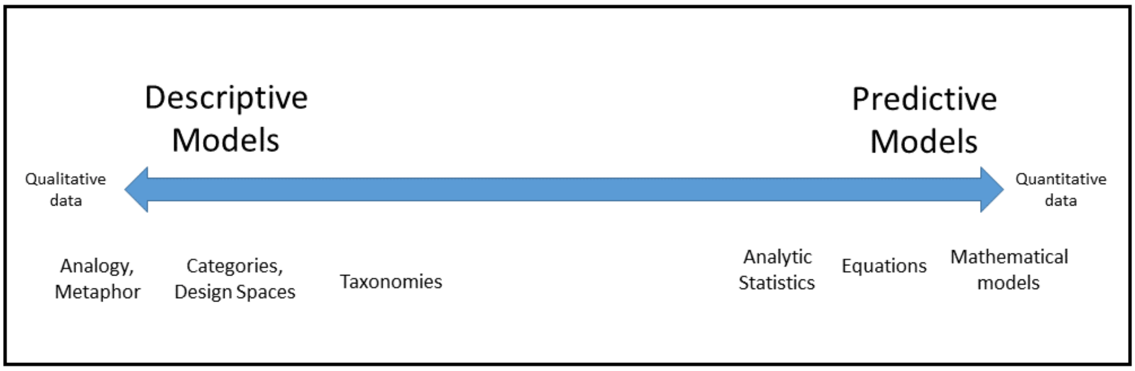


Figure 3. A Continuum of Models, adapted from [11] with descriptive and predictive models at opposite ends of a spectrum going from HCI outputs mostly based on qualitative data to the ones mostly based in quantitative data.

PCMs are especially relevant for the study of IVEs because they demand for a more empirical approach, requiring the development of highly controlled user-testing procedures and outputting quantitative user performance data. At a time where IVEs are becoming serious contenders for turning into the next pervasive UI, there are still some perceptual phenomena that need to be fully understood to allow for the creation of better user-experiences. Moreover, there is a lack of standardized procedures capable of being used to evaluate these ICSs. Thus, in the scope of this thesis, we intended to generate predictive models of user perception and performance in IVEs by developing protocols (and their implementation) for assessment of human perception and action in IVEs. These assessments will allow the collection of data on users' perception and performance, and comparison with existing *models of human perception and performance*. The contrast between obtained results and the predictions of models of human perception and performance provides development/adaptation guidelines for the systems under evaluation. Following the recommendations of ISO9241 – *Ergonomics of human-system interactions* – developers should adopt an iterative approach in order to use such tasks for evaluation and development of ICSs (see **Figure 4**) [13]. Such approach, by its nature, emphasizes the need to understand user's capabilities and limitations in order to guide development and establish the minimum requirements for highly immersive virtual environments. We consider that this approach might be highly valuable for creating new IVEs in different areas of application such as gaming, data visualization and manipulation tools, and even virtual set-ups for controlled experimentation on human perception and behavior.

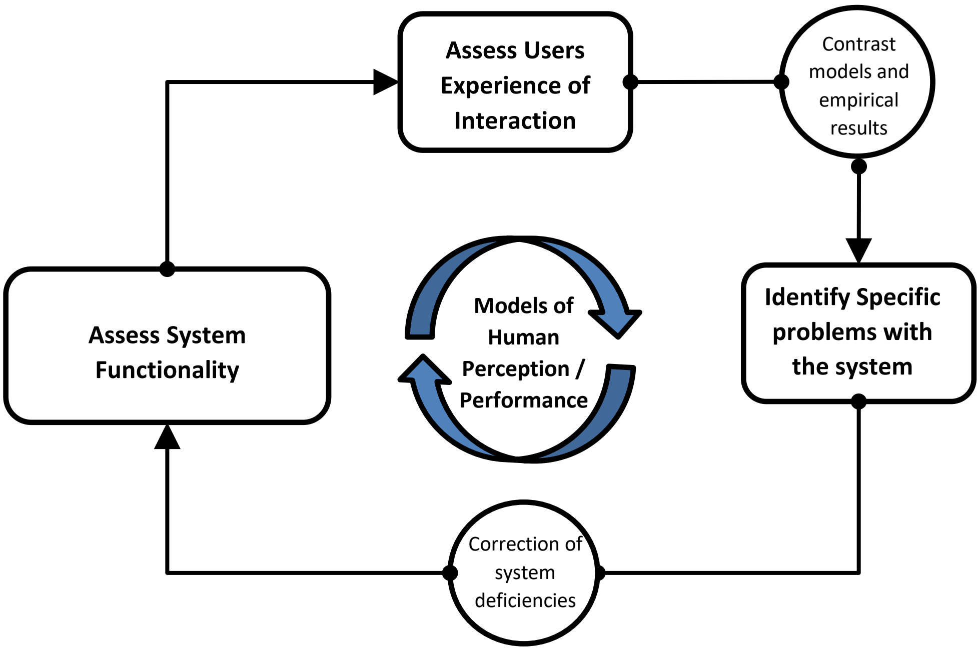

Assess Users Experience of Interaction
Contrast models and empirical results
Identify Specific problems with the system
Models of Human Perception / Performance
Assess System Functionality
Correction of system deficiencies

Figure 4. Diagram of the evaluation/development iterative process (adapted from [10]). Crucial to this process is the contrast between empirical results and models of human perception / performance. The goal of this iterative process is to obtain empirical results that are in line with established models of human perception or performance.

DCMs, on the other side, by offering a comprehensive description of all the interaction process, are particularly important during the study and development of safety-critical devices. This type of interaction models can assist in the development of fault-tolerant devices by helping researchers understand how UI and use error are related. Thus, in the scope of this thesis, we intended to generate better descriptive models capable of covering all UI faulty modes of medical devices and accounting for different type of users or context of use. Moreover, by analyzing the most common root-causes for use error, we will use DCMs to guide the extraction of design principles capable of lowering the probability of faulty user-machine interactions.

Throughout this thesis, we will be presenting different controlled-assessment scenarios for evaluation of users' perception and performance when using interactive computing systems. Our use cases will span from the accurate implementation of aural-visual interaction in a multimodal CAVE-like system, to the study of use-error in medical devices and the development of a Simulation-Based Medical Education (SBME) IVE for training of user interaction with safety-critical devices.

1.4. Thesis Overview

In **Part II** of this thesis, we will begin by presenting the theoretical background on IVEs and PCMs (**Chapter 2**). A state of the art and a description of the limitations of current IVEs will be provided and we will describe how PCMs have contributed for the development of ICS in the past, as well as making the case that they can also contribute for the development of improved IVEs. In **Chapter 3**, we will present the application of empirical user evaluation methodologies in a diverse set of psychophysical experiments on audio-visual integration, visual distance perception, and auditory localization (CAVE use case). Direct comparison of users' performance with existing models of human perception and performance will serve as benchmarks in order to propose empirical-based adaptation to the IVEs, suggest IVEs development guidelines, or to aid in the selection of technology to be used in the IVE implementation. An evaluation framework compiles and documents the several empirical evaluation methods on software capable of being deployed to different IVE platforms. In its essence, this framework allows for user evaluation and, consequently, informed adaptations of the IVEs that have the potential of greatly increasing immersion.

Following a similar structure, in **Part III**, we will begin by presenting the history of safety-critical medical devices and by reviewing DCMs that are especially concerned with the problematic of use-error (**Chapter 4**). In **Chapter 5**, we will demonstrate the role of analytical evaluation methodologies while studying in detail use-error in safety-critical systems (medical devices use case). The outputs of this methodology will be used to develop a new use-error taxonomy that may serve as the basis to derive design guidelines for safer safety-critical UIs.

Finally, in **Part IV**, we will summarize the main outputs of this thesis and frame their importance in the context of the future of ICSs (**Chapter 6**).

Part II

PCMs and Immersive Virtual Environments

2. THEORETICAL BACKGROUND ON IVEs AND PCMs

"We must design for the way people behave,
not for how we would wish them to behave."
Don Norman, *Living with Complexity*

In this chapter, we will present the state-of-the-art of IVEs and discuss some of its usability issues. We will also review the history of interaction-modeling methodologies based on Predictive Cognitive Models and show how these methodologies have been contributing for the development of improved ICSs.

2.1. What are Immersive Virtual Environments?

"Immersive" is a characteristic that we can rightfully attribute to a breathtaking landscape, to the prose of a classic novel, or to a brutal scene displayed on a captivating painting. Lately, however, the term has been widely applied to the description of technological endeavors, such as computer-generated environments. In fact, the term "immersive" has been applied to computer-generated environments so often, that they deserved a direct reference on its Merriam-Webster entry. This dictionary defines the adjective *Immersive* as the ability to involve or absorb in *something*, this something being an activity or a real or *artificial environment* [14]. The novelty in this definition of the term *immersive* is in the reference to artificial environments, strategically included to reflect the extensive use of this adjective in the computer graphics and virtual reality industry.

On the other hand, the term "virtual reality" has a fully dedicated entry on the Merriam-Webster dictionary, which is focused entirely on the technological aspect of the term: "*an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment*" [15].

Because both terms in isolation fail to capture the essence of the main subject of this thesis, we combined both concepts and propose the following definition for *Immersive Virtual Environment (IVE)*:

> *A computer-generated environment that, upon receiving input from users' actions, conveys the appropriate stimulation to different users' senses in such a way that naturalistic interactions with virtual elements become possible.*

Breaking down this definition, we can agree that there is a first part to it that is close to the one provided from the Merriam-Webster for "virtual reality", and this is intended in order to establish the technological requirements of an IVE. This part states that the environment is *computer-generated* and it is responsive to users' actions, thus it requires sensors that capture human behavior and actuators that render appropriate sensorial stimulation. On the other hand, the second part completes the Merriam-Webster definition for "virtual reality" by adding the reference to these environments' immersive property by stating that they should allow for *naturalistic interactions* with virtual elements. While, normally, users do not lose the ability to discern reality from *virtuality* when interacting with an IVE[6] (nevertheless, see [16]), quite often it is possible to capture realistic reactions to virtual elements [17], [18]. These realistic reactions are only possible if, indeed, the user becomes absorbed, immersed in the virtual world. Thus, we can say that IVEs have the ability to create, in its users, the illusion of being in a place other than where they actually are, or of having a coherent interaction with objects that do not exist in the real world – normally referred to as a *feeling of presence* [19].

Feeling of presence is an important subjective experience conveyed by the use of IVEs which, in the words of Sheridan ([20], p. 120), can be summarized as: "…*a sense of being physically present with virtual objects*". The feeling of presence renders the virtual environment credible and greatly contributes for an acceptable interaction with virtual objects. In order to convey this type of sensorial experience there are some features that an IVE should present:

1- **Unnoticeable hardware:** An IVE should aim to convey the most naturalistic interaction possible. In this sense, obstructive hardware can be a problem as it functions as a constant reminder of the simulation limitations. A *Head Mounted Display* (HMD) might weight around 470 grams[7], while reading glasses weight around 100 grams. The more the hardware of an immersive system goes unnoticed (consider not only the weight of the display and peripheral equipment, but also the existence of cables, markers, and similar components) the better immersion it will convey (see e.g. [21] and [22]);

[6] Nevertheless, people are actively trying to create set-ups where they can lose discernment of reality while using VR apparatus. In March of 2019, Jak Wilmot, the co-founder of Atlanta-based VR content studioDisrupt VR, spent a total of 168 consecutive hours (1 week) in an VR headset, either seeing VR environments or pass-through computer generated images of its living room. The original report of this experience can be read here: https://futurism.com/guy-week-vr-headset/

[7] 470 grams is the spec weight for both the Oculus Rift Consumer Version and the HTC Vive. The Sony HMZ-T3W specifications claim a weight of 320 grams, although battery, processor, and cable are not accounted for.

2- **Real-time update of the immersive environment:** a quite important feature of IVEs is stimulation update based on user's position and orientation tracking. The immersive capacity of a system will increase greatly if the image and/or the sound conveyed are dependent on the head movement of the user (and, obviously, congruent with the physical changes expected in the real world). This is, as we will see, still one of the mains sources of problems in today's IVEs and one of the areas that can benefit the most from the capacity to adequately calibrate real-time responses against what is expected by users;

3- **Multimodal stimulation:** The feeling of presence and consequent immersion appears to increase with the number of sensorial modalities that an immersive system is able to stimulate [23]. Theoretically, if the purpose is to convey *full immersion*, all the five senses should be stimulated using audiovisual *Virtual Reality* (VR) in combination with haptic stimulation and force feedback, smell, and taste replication. Nowadays there is no system that satisfactorily conveys a sensation of full immersion, nevertheless, multimodal immersive environments are a long existing reality and are regarded as more immersive than unimodal immersive systems;

4- **User-environment interactivity:** Interactivity between user and the three-dimensional representations in the virtual world is paramount for an increasing sensation of immersion. Participants who were allowed to interact (e.g. move, rotate, catch) with stimuli of a virtual environment, report a higher feeling of immersion than passive observers [19].

Immersion and feeling of presence in a virtual environment are inseparable and positively correlated with the measured quality of the user experience in an IVE [24]. Thus, today's developers aim to come out with ever increasingly immersive virtual reality experiences and any technological advancements in that sense puts them one-step ahead of the competition on a multi-billion dollar market[8]. Nevertheless, while technological developments on the four crucial properties for presence (unnoticeable hardware; real-time update; multimodal stimulation; interactivity) are constantly reaching new levels of performance, little progress has been done in order to understand the fundamental part of the immersion process: *the user*. In the next section, we explain why the current main challenge, when developing IVEs,

[8] Augmented Reality and Virtual Reality Market are valued at USD 4.21 and 5.12 Billion dollars respectively, in 2017. Both are set to grow at a CARG above 30% up to 2023. The world's biggest investment fund, the SoftBank's Vision Fund managed by Masayoshi Son, has one of its highest investments on *Improbable*, a British company aimed at creating virtual environments indistinguishable from real world. See *The Economist* May 12th 2018 edition – "The Son Kingdom" report.

is not a technological one, but instead it is precisely a matter of understanding the interaction between these systems and the human element.

2.2. What are the problems with IVEs and how can we solve them?

IVEs have the core goal of conveying the illusion of presence of (or in) something that, in fact, does not physically exists [24]. The systematic study and development of computerized IVEs can be traced to the early works of Ivan Sutherland and his student Bob Sproull that developed what has become known as the first *Head Mounted Display* (HMD), presented in 1968 (**Figure 5**). This was a system so heavy that had to be suspended from the ceiling thus inspiring its name, *The Sword of Damocles*[9]. This HMD was the first to use computer-generated visual stimuli capable of adapting to user head position. By today's standards, it was a very primitive system of *Augmented Reality* (AR), both in terms of interface design as in terms of graphical representation of the virtualized stimuli. It was only capable of comprising simple wireframe cubic rooms in the binocular display, nevertheless the perspective of the room shown to the user was already dependent on the user's position. The Sword of Damocles was a typical see-trough AR display, because images were conveyed on half silvered prisms, enabling a view of both the synthetic imagery and the real world surroundings (see definition of *mixed reality*, **Section 2.3**).

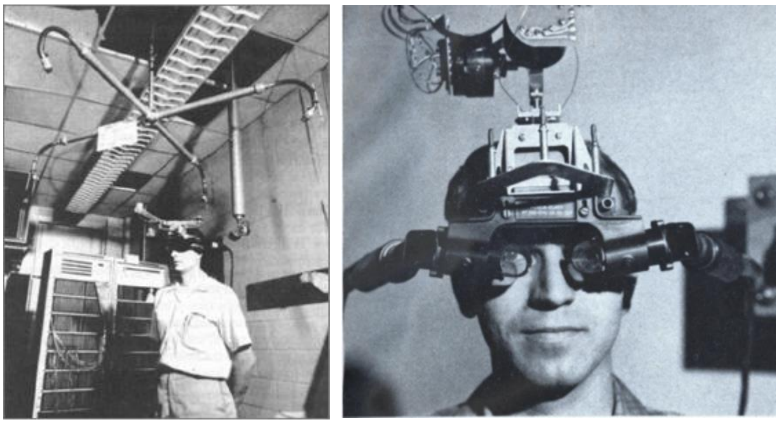


Figure 5. The Sword of Damocles, a see-through Augmented Reality device.

Ever since this computer science's milestone, IVEs have gained crescent notoriety, alternating periods of high notoriety with periods of apparent vanishing from the scientific and academic debate [25]. Recent theoretical and technological developments have contributed for some consolidation of IVEs as a

[9] The name is a reference to an anecdotic passage in Cicero's *Disputations*, where the roman orator tells the story of Damocles a courtier of Dionysius II of Syracuse. Damocles once said to the king that he, as a great man of power, wisdom, and luxury was a truly fortunate man. Therefore, hearing this, the king offers to switch places with Damocles on the condition that he had to feel the same endangerment to his life as the king daily felt. To simulate this sense of endangerment, a huge sword was hanged, held at the ceiling by a single hair of a horse's tail, right above the throne and directly facing the head of the new throne owner, Damocles. After a short period of time, Damocles ended begging Dionysius to switch places again.

consumer good technology and for the emergence of an increasingly number of VR and AR applications in the most varied fields, such as: science and data visualization, medical and surgical devices, telepresence systems, flight and driving simulators, therapy tools, digital arts, cinema, and videogames. Because of its inherent potential to directly interact with the human senses, immersive environments that make use of VR or AR have since long been regarded as in line to become the next predominant human-computer interface. Carolina Cruz-Neira and her team suggested this possibility following the enthusiasm with the first Cave Automatic Virtual Environment (CAVE) system [26] (see **Section 2.6**), and more recently [27] and [28] referred immersive operative systems like the HoloLens, as in line for the next revolution regarding new concepts of user-interface (UI). However, in order to turn IVEs into a serious contender for the next predominant interface paradigm some current technologic and human factors limitations have to be overcome and, additionally, development approaches more focused on the analysis of human perception and action ought to be pursued.

As soon as 1965, Sutherland had already outlined the ultimate challenge for IVEs developers in what has become known as the *Sutherland's challenge*:

> "*Make that (virtual) world look real, sound real, feel real, and respond realistically to the viewer's actions*"[10] [29].

More than fifty years later, and in a time where IVEs are becoming of widespread use, users still complain about their odd perceptual effects and performance deficits. These deficiencies affect all four factors involved on the triggering of the feeling of presence, mentioned in **Section 2.1**:

1 - **Unnoticeable hardware:** Today we already have wireless IVEs in two main forms: Smartphone VR Headsets (such as the Samsung Gear VR) and projection-based systems like the CAVE. It is true that in both systems users are not restricted by cable connections, however, while for Smartphone VR Headsets VR applications are still confined to passive-viewer and low quality video streaming, CAVE systems allow for user interactions with high-resolution virtual objects in large projection surfaces (room-like). Unfortunately, the CAVE system is not considered an electronic consumer good due to the high costs involved in assembling such

[10] Sutherland's conceptualization of the *ultimate display* also included some incursions in more extravagant predictions, such as: "*The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such room would be fatal*" [29].

virtual environment systems[11]. In terms of cost and quality, a middle-range solution is the PC-based premium VR headsets, commonly referred to as HMDs (such as Oculus Rift, HTC Vive, and the PlayStation VR). However, currently these systems still require cable connection[12] and the weight and size of these headsets keep raising complains from the users [24], [30]. In order to achieve wireless professional level HMD-based IVEs, developers will have to wait for 5G connection promises of wireless high throughput and low-latency connections to hold true [31].

2 - **Real-time update:** Handling the delay between tracking the user's head position and congruent change in the projection of a 3D image or of an auralized (3D) sound is one of the major challenges in the development of IVEs [32]. This type of delay is commonly referred to as *end-to-end delay* [33], and it is the consequence of the accumulated latency in several individual components, including tracking devices, signal processing and communication, and output devices. Consequences of end-to-end latency can range from output errors in terms of the visual perspective or 3D sound position, to more serious consequences as simulator motion sickness [34] [35]. End-to-end latency in IVEs results in temporal-spatial contradictory sensorial inputs to the user's visual and vestibular systems. While the vestibular system might indicate that the user is tilting his head downwards 45 degrees, the visual render might still be displaying a 0 degrees perspective due to update lag resulting from end-to-end delay. These visual-vestibular incongruences are the main etiological factor of motion sickness [36] and some studies indicate that the delays in linear visual oscillations (also known as *heave* motion) are particularly prone to cause motion sickness [37]. Some authors state that end-to-end delay should be below 15-20 ms in order to prevent simulator motion sickness [31], nevertheless, while some manufacturers already claim performance levels around these figures, controlled measurements report otherwise [38]. This is even more problematic in wireless collaborative IVEs, since current 4G communication involves a minimum of 25ms latency in controlled ideal operation conditions [31].

3 - Multimodal stimulation: Truly multimodal IVEs, developed as such since the conception phase, are still scarce. Despite few exceptions from the past (such as the Morton Heilig's

[11] A Christie Mirage DLP projector can surpass the price tag of 50 000€, depending on the model.

[12] Recently, speculation has surface about an Oculus Rift new generation headset codenamed Santa Cruz. Reports are stating that this model will be wireless, possibly taking advantage of 5G connection. The company has stressed that this is still a prototype and not a commercial product. See: https://www.theverge.com/2017/10/12/16463844/oculus-santa-cruz-standalone-headset-prototype-hands-on

Sensorama (see [39]) and the visuo-tactile VR system outputted from the European project PURE-FORM [40]), IVEs are predominantly unimodal. Moreover, multimodal IVEs are usually restricted to the actuation on two sensorial modalities – such as visual IVEs coupled with motion actuation (e.g. driving simulators), or audiovisual IVEs where stereoscopic imagery is coupled with spatialized audio. Current platforms of IVE development (such as Unity, Blender or Odeon) do not offer the possibility of developing with the same level of control for more than one modality. They are either fully focused on the development of the visual IVE (Unity and Blender) while offering comparatively poor add-ons for spatialized audio, or they are exclusive for acoustic modeling (Odeon). Moreover, users' different sensorial modality channels operate with different timing constrains (see **Section 3.1**), which coupled with the end-to-end latency problem renders the synchronization of multimodal IVEs quite a challenging task.

4 - **User environment interactivity:** Seamless and convincing interaction with virtual objects is still one of the main challenges in IVEs development. This is a two-part problem where *seamless* and *convincing* are utterly difficult to combine in the same interaction actuator. On one hand, interaction is hampered due to the requirement to use peculiar peripherals (such as the HTC Vive wireless controller or the Oculus Touch VR controller), which require a learning process and prevent the use from performing natural movements (i.e., grabbing, catching, and pushing virtual objects) due to handling requirements of the controller. Nevertheless, seamless means of interaction are started to be employed mainly by using optic motion capture systems such as the Vicon™ and Kinect™[13]. On the other hand, convincing interaction will require technologically complex (and not that seamless) tactile and haptic actuators, so the user can both interact and sense the physicality of the virtual object. Today's haptic-actuators are mainly relying on vibration and force-feedback, which still require contact, or at least close proximity with fairly obstructive pieces of equipment [41] (see [42] for one of the less obstructive solutions). Contactless haptic actuators as ultrasound and air vortices are starting to be explored [43], however they are still very limited regarding their ability to simulate complex stimuli features such as size, shape, and texture of an object.

[13] Although Kinect started as an prized success – was on the Guinness Book of World Records as the fastest-selling consumer electronics device with 8 million units in 60 days – Microsoft recently announced it will stop its development and production on the grounds of diminishing financial returns (see https://www.popularmechanics.com/technology/gadgets/news/a28783/microsoft-ends-production-of-kinect-attachment/).

As we can understand from the examples above, there are still several problematic issues with today's IVEs. While some of these problems translate in almost purely technical challenges (reducing the equipment weight, for instance, or develop integrated virtualization platforms for multimodal IVEs), others will require user studies and fundamental research in Human Perception in order to support the development of technical solutions. Take, as an example, the above-mentioned difficulty in synchronizing the output of stimuli from different sensorial modalities. To resolve this problem, an in-depth knowledge about the human perceptual mechanisms that allow for synchrony perception is required. For instance, it is known that humans still perceive as synchronous an audio-visual stimulus where sound lags image for some milliseconds. This capability has been attributed to the existence of a *Window of Temporal Integration* (WTI) [44]. Nevertheless, we still need to know if this tolerance to lagging sound allows accommodating the typical audio-visual mismatches due to end-to-end delays found in the current IVEs. Moreover, we need to understand if this WTI is affected by some other characteristics of the IVEs, such as the distance at which the virtual stimulus is presented to the user. We will be looking in more detail into this problematic in **Section 3.1**.

Because models of human perception and performance are the end-result of this empirical approach, as well as the benchmark of the evaluation/development iterative process described in **Figure 4**, in the following section, we will explore in more detail how these models can guide us in the development of increasingly immersive virtual environments.

2.3. On Understanding and Modeling Perception and Performance for the Development of IVEs

Accurate predictions about how users perceive and interact with a computerized environment are of the foremost importance in the design of computer systems that emphasize usefulness and usability. Furthermore, any attempt at developing an interactive system should put the human, the *user*, in a central position that defines all the subsequent discussion and design [10] [13]. Knowing and understanding the perceptual mechanisms that allow users to perceive the natural world should be indispensable for the design of any satisfying interactive environment which makes use of audiovisual virtual or augmented reality.

Computer generated immersive environments are broadly classified into two different categories: VR environments and AR environments. The distinction between these two is not a procedural or technical one; rather it is more of a performance-based distinction. The *Reality-Virtuality (RV) continuum* of Milgram and Colquhoun [45] has been widely used to clarify this distinction and provides a good theoretical framework to derive a taxonomy on immersive environments. In the reality-virtually continuum the fundamental distinction is between *Real Environments* and *Virtual Environments*, that are located on opposite ends of the continuum (see **Figure 6**).

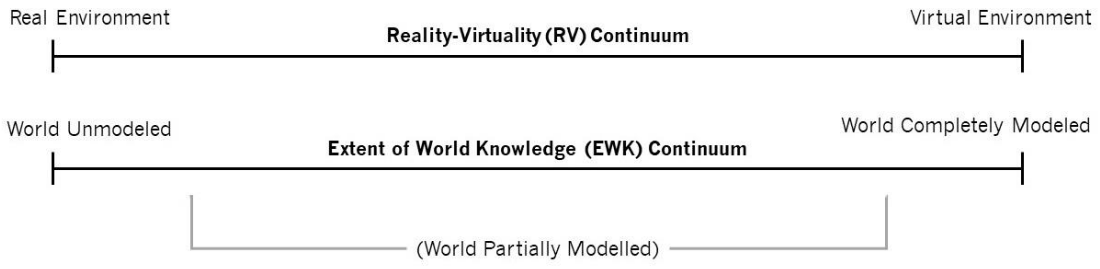

Real Environment
Virtual Environment
Reality-Virtuality (RV) Continuum
World Unmodeled
World Completely Modeled
Extent of World Knowledge (EWK) Continuum
(World Partially Modelled)

Figure 6. Reality-Virtuality Continuum (adapted from [45]).

The location of any computerized immersive environment along this continuum coincides with its location along a parallel *Extent of World Knowledge (EWK) continuum*, highlighting in this way, the importance of knowing about both the physical world and the human perceptual mechanisms that govern the perception of these physical signals. As depicted in **Figure 7**, on the right end of the R-V continuum are virtual environments (which need to be completely modeled in order to be rendered) implying, in this sense, full knowledge about the physical properties of the world and how they are perceived by humans. However, if we increase the granularity of the R-V continuum we can identify *Augmented Reality* (AR), as an environment where more than half of the scenario is real environment with no modeling whatsoever; and *Augmented Virtuality* (AV), as an environment where more than half of the scenario is virtual environment. These two types of intermediate stages are sometimes referred as *Mixed Reality* (see **Figure 7**).

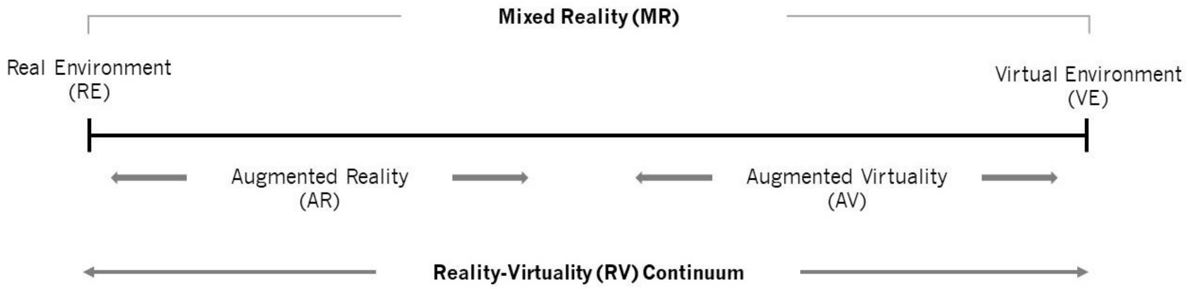
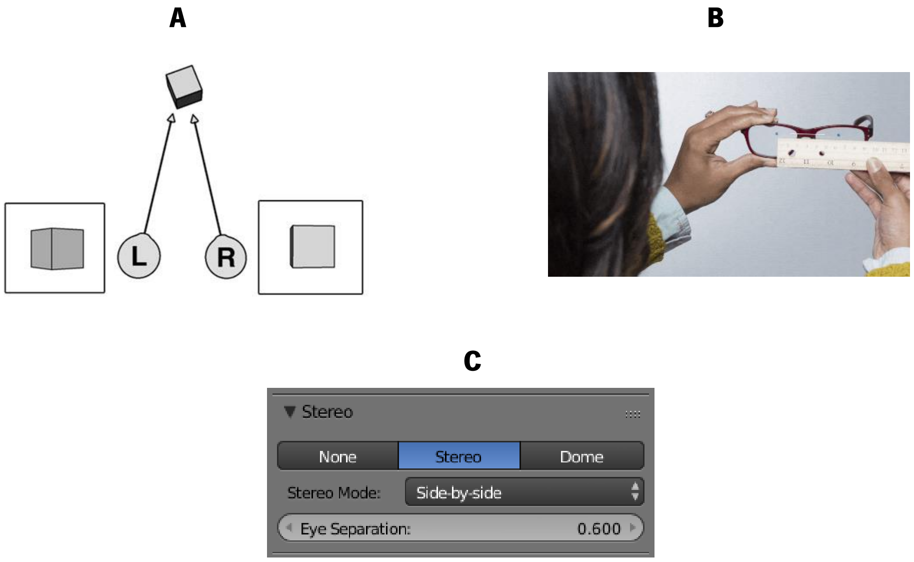


Figure 7. Reality-Virtuality Continuum, AR, and Augmented Virtuality (adapted from [45]).

A clear example of the close relation between the EWK continuum and the R-V continuum is the simulation of binocular depth cues in order to provide visual depth perception. The technique of *Stereoscopy* can be traced back to the first half of the nineteenth century [24], although initial demonstrations of the effect were developed when the principles of retinal or binocular disparity were discovered in the XVI century [46]. In stereoscopy, depth is simulated based on the fact that we naturally process two slightly different images of the same visual scene – one for each eye – and the bigger the offset between the two images the closer the object is perceived (see **Figure 8**). The differences between the images that each eye processes can be calculated based on the person's eye separation and thus, binocular depth cues can be easily simulated if we control the image that each eye is seeing and if that image is adjusted for the position of the correspondent eye.



Stereo Mode: Side-by-side

Eye Separation: 0.600

Figure 8. Retinal image offset due to binocular disparity (A). Measurement of pupillary distance (B) performed after fixating on an object around 6 meters away and by marking each eye fixation with a marker. Use of the pupillary distance as a parameter to define the geometry of stereoscopic scene in some virtualization platforms such as in the Game mode of the Blender software [47] (C).

The relevance of human' studies for the development of IVEs is latent in the parallelization between the R-V and the EWK continuum. The most important principle in the design of immersive environments states that the environment should convey an accurately replication of the geometric and temporal characteristic of the real world. This does not mean that computerized environments must model precisely all physical attributes of a visual or an auditory real world's scenario – in fact this is, and it is expectable that will remain being, technologically impracticable. What this principle really means is that the perception that a user has in a VR or AR environment, should be *quantitatively indistinguishable* from a correspondent scene perception in the real world. In the words of Machover and Tice in their seminal paper on Virtual Reality [48]:

> "*...the quality of the experience is crucial... the 'reality' must both react to the human participants in physically and perceptually appropriated ways, and to conform to their personal cognitive representations of the microworld in which they are engrossed. The experience does not necessarily have to be realistic – just consistent*" (p.15).

Helping to measure performance in order to convey a consistent experience is where the discipline of HCI can play a key role. Applying HCI methodologies to the development and evaluation of immersive environments becomes an issue of the outmost importance, primarily when we think about the current lack of knowledge about:

(1) *human perceptual mechanisms* – e.g., the mechanisms governing natural audio-visual asynchronies, the binding of perceptual cues to yield visual and auditory depth perception, and audiovisual recalibration phenomena;

(2) *human performance in IVEs* – e.g., realistic interaction with virtual elements, auditory and visual depth perception and performance validity on high fidelity simulators;

Knowledge about these two dimensions (human perception and human performance) is central to replicate the natural tridimensional world interaction conditions in IVEs. Thus, in the IVEs context, the use of HCI interaction-modeling methodologies and empirical research techniques can both give us some insight about the human perceptual mechanisms that are crucial for the design of a highly immersive VR or AR environment and, at the same time, allow us to accurately evaluate human performance of an IVE user and compare it with models of human performance in the real world. This comparative evaluation had widespread used in the development of PCMs that have boosted HCI influence in the design of digital interfaces. The same approach should be replicated for the development of IVEs, particularly if we expect them to become a credible and widespread UI option.

Of particular importance for user studies in IVEs, is the application of Psychophysical Experimentation methods, an approach borrowed from cognitive sciences that allows for clear and quantifiable insights on human perception and action.

2.4. What is Psychophysics and how can it change the way we interact with computers

We argued that the development of increasingly immersive virtual environments is dependent on the knowledge about both the characteristics of the physical world and how humans perceive and interact with physical events. In this sense, we consider that being capable of measuring human perception and performance is a fundamental part of developing a highly immersive environment.

Vicki Bruce, an experimental psychologist, defined Psychophysics as: "*the analysis of perceptual processes by studying the effect on a subject's experience or behavior of systematically varying the properties of a stimulus along one or more physical dimensions*" [49]. Since the first half of the 19th century (starting with the works of Ernst Weber and, later, Gustave Fechner) psychophysics, as an established scientific field, tried to understand and describe the perceptual mechanisms responsible for turning some physical phenomena into psychological phenomena. Psychophysicists usually develop experimental tasks in which participants are presented with stimuli with several physical properties that can be precisely measured (i.e. brightness, contrast, angular size, intensity, just to name a few…). In this way, and by studying how varying some physical properties of the stimuli affects human perception, it is possible to determine what are the critical features of a physical signal that convey a certain perceptual phenomenon [50]. Traditionally, the research in psychophysics tries to discover three characteristics of a given perceptual phenomenon: *absolute thresholds* (signal detection), *differential thresholds* (signal discrimination), and *magnitude estimations*. Any perceptual event that relies on one of these three mechanisms, independently of the sensory modality involved, is an object of study to psychophysics.

A good example of how psychophysics can help shape the technology development is the case of research on human auditory acuity and development of listening devices. The science of measuring the human auditory acuity is since long well established [51], but its repercussions still affect the way listening devices are designed today. Using signal detection tasks, a researcher would present, through a *staircase method*, tones of a specific frequency but varying in intensity that would require a response from the listener signalizing that he could hear that sound (e.g. by pressing a button when hearing). Repeatedly decreasing the sound intensity when participants can hear, and increasing when participants cannot, allows the research to define a *limen/threshold* for that given tone frequency. What the psychophysics research showed is that humans have a very different auditory acuity to different sound frequencies

(**Figure 9**). Humans are more sensitive to the range of frequencies around 200Hz-7kHz, curiously the range of frequencies spanning human speech, and are worst handling low and very high frequencies.

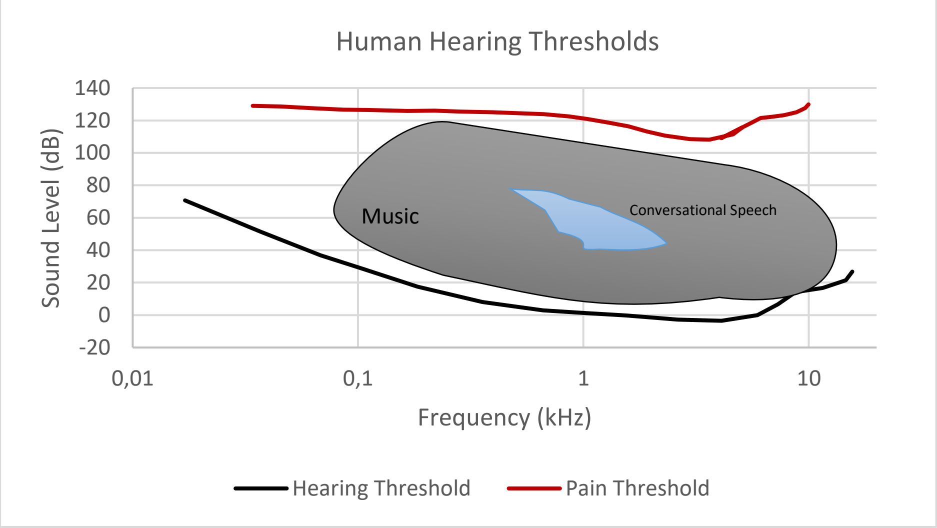


Figure 9. Representation of the human auditory acuity in decibels per frequency. The area between the black and the red line corresponds to the range of audible sounds. Any sound below the black line is typically inaudible and every sound above the red line is potentially harmful for the human hearing system. The grey area shows the range of frequencies and sound levels normally applied in music, while the blue area shows the range of conversational speech. Adapted from [50].

The model of the human hearing threshold was soon adopted as an indication for listening devices developers, which have devised three main types of headphones (see **Figure 10**).

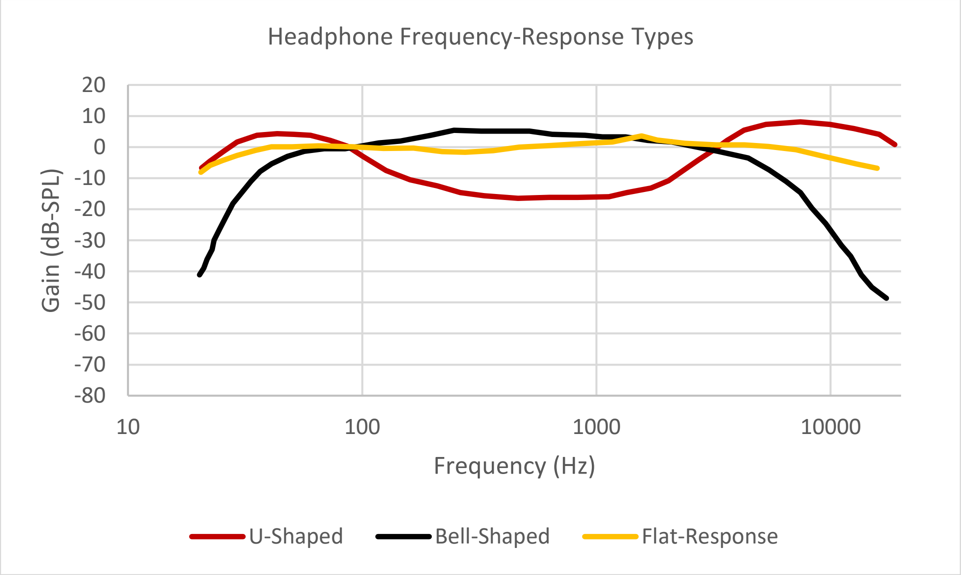


Figure 10. Frequency response of different types of headphones.

These types are divided according to the purpose of use and are mainly affected by the human hearing threshold. Thus, if the goal is to have a high-fidelity headphone that accurately reproduces what was captured, we should choose headphones with a nearly uniform frequency response (yellow – flat frequency response). However, despite being faithful to what was captured, most of the time these headphones are not the best ones for enjoying music, for instance. In this case it might be preferable to have U-Shape frequency response headphones (red – U-shaped frequency response), because it will emphasize the bass and the high frequencies thus ensuring that the different notes from different instruments are well perceived and listened to. On the other hand, the goal for our listening device might be to enhance human speech, like on a mobile or tele-conference equipment. Thus, if we want to use these devices with the primary goal of communicating with other persons, we might want to enhance the frequencies in the mid-range (black – bell-shaped frequency response) and, in fact, that is what some microphones and cell-phones headsets do – they filter high and low frequencies when capturing sound and boost frequencies in the mid-range when transmitting sound. Some software already applies these principles in smart equalizing tools (see for instance the *Sonible software*), to better adapt the result of sound mixture process to purpose of the final sound experience. As we can see, the accurate definition of the human hearing threshold – which is by itself a perceptual type PCM – opened the doors for a range of different applications in the acoustic industry.

This type of applications allow us to easily understand the role of PCMs in the development of more usable UIs. With the formalization of the knowledge about the users' perceptual or cognitive processes, developers gain a model that can directly affect and guide the design of technological outputs without the need to run lengthy, and sometimes costly, user experiments. Nevertheless, it is important to understand that, in order to get to a powerful PCM (in the sense that it conveys accurate predictions of user behavior) with real impact on the design of UI, HCI researchers needed to first conduct extensive user experiments to understand the phenomenon. In the next section, we will review some influential examples on the development of new PCMs and present their impact in areas such as software engineering, product design and even cognitive sciences.

2.5. Human Predictive Cognitive Models

The final goal of PCMs is the prediction of human performance, but in order to achieve this goal the models should give account of how information is processed and what is the role of human perception, memory, attention, and motor skill in the outcome. Next, we will present six examples of PCMs that made an everlasting impact in the discipline of HCI: the Fitts' Law; the Hick-Hyman Law; the Keystroke-Level

Model, the Model Human Processor, the ACT-R, and the EPIC model. We will look at these examples in terms of (1) how they were developed and (2) what impact they brought to the development of new UIs.

Fitts' Law

A pioneering work that fostered the use of PCMs in guiding the design and evaluation of computerized interfaces was the work of experimental psychologist Paul Fitts. In two highly influential papers – one in 1954 [52] and another in 1964 [53] – Fitts presented an empirically grounded model to explain human motor behavior. As an enthusiast of *Information Theory* and an advocate of its use to explain sensory, perceptual, and perceptual-motor human behavior, Fitts draw an analogy from the well-known Shannon's Theorem 17[14] [54] to predict human performance on a tapping task – a task requiring similar perception-action skills to those applied in today's pen-based user interfaces.

To test the adequacy of information theories to motor behavior, Fitts developed a highly controlled task, named *reciprocal tapping*, in which participants had to tap two rectangular metal plates alternatively with a stylus. Movement tolerance (accuracy needed to tap the metal plate) and movement amplitude (space to hover between the two plates) were controlled for by presenting the participants with different width plates at different distances between them [52] (see **Figure 11**).

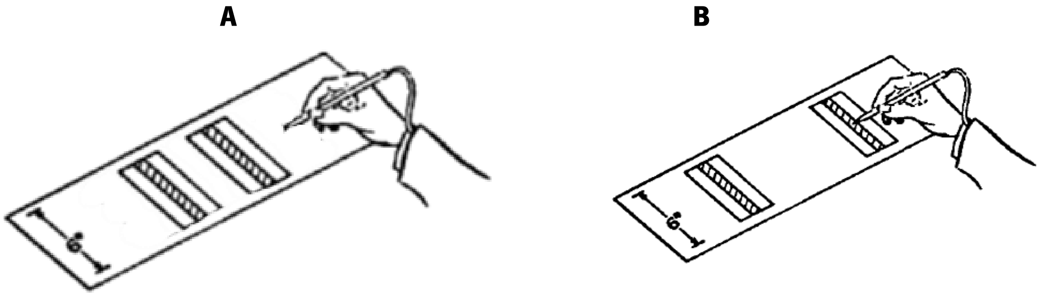


Figure 11. Two conditions of the reciprocal tapping experiment, with different movement tolerance and amplitude. **Panel A** represents an easier condition, because the distance to be covered between the two plates is lower and the tapping surface is bigger. Fitts' law predicts shorter movement times for this condition, while predicting longer movement times for the condition represented in **Panel B**, where the distance to be covered is higher and the tapping surface is shorter.

The apparatus was composed by two surfaces that should be taped alternately, each one composed of three plates – the target plate, the overshoot plate (out-most portion of the surfaces), and the undershoot plate (inner portion of the surfaces). These "error plates" where wide enough to capture all over and undershoots. In the original experiment, Fitts varied the center-to-center distance of the target

[14] Theorem 17 states that the capacity *C* (in bits/s) of a communication channel with bandwidth *B* (Hz) can be described as $C = B\,log_2\left(\frac{S}{N}+1\right)$, where *S* is the signal power and *N* is the noise power.

plates (movement amplitude) in four different conditions: 5.08, 10.16, 20.32, and 40.64 cm[15]; while the width of the target plate was either 0.64, 1.27, 2.54, and 5.08 cm[16] in a total of 16 different possible arrangements. Thus, the most difficult task condition was the one where target plates distance 40.64 cm from each other (greatest movement amplitude) and were only 0.64 cm wide (smallest tapping surface), while in the easiest condition they were only 5.08 cm apart while being 5.08 cm wide. One can easily order the different 16 arrangements in an increasing level of difficulty; nevertheless, Fitts suggested that the *Index of Difficulty* (ID) of the task followed a function of the Information Processing type, which is:

Equation 1. $ID = log_2 \left(\frac{2A}{W}\right)$*, where A is movement amplitude and W is target width, in centimeters.*

Thus, as information capacity improves with increasing signal-to-noise ratio, similarly, the difficulty index of a motor task also increases as the ratio between the distance to cover and the size of the target increases. Stated in this way, this relationship seems quite common sense. However, what is more impressive is that with the help of empirical data Fitts was able to show that this relation holds true for several motor tasks and that accurate prediction of task performance time is possible. **Table 1** shows a comparison between predicted and observed time for each task condition (from a set of 16 participants).

Table 1. Different task conditions in [52] and respective predicted (using **Equation 1**) and observed time-to-completion.

W	***A***	***ID***	***Real Time (s)***	***Predicted Time (s)***
0,64	5,08	4	0,392	0,391
0,64	10,16	5	0,484	0,485
0,64	20,32	6	0,580	0,580
0,64	40,64	7	0,731	0,675
1,27	5,08	3	0,281	0,297
1,27	10,16	4	0,372	0,392
1,27	20,32	5	0,469	0,487
1,27	40,64	6	0,595	0,581
2,54	5,08	2	0,212	0,202
2,54	10,16	3	0,260	0,297
2,54	20,32	4	0,357	0,392
2,54	40,64	5	0,481	0,487
5,08	5,08	1	0,180	0,107
5,08	10,16	2	0,203	0,202
5,08	20,32	3	0,279	0,297
5,08	40,64	4	0,388	0,392
Mean Time			**0,392**	**0,391**

[15] Originally: 2, 4, 8, and 16 inches respectively. See [52].

[16] Originally: .25, .5, 1.0, and 2.0 inches respectively. See [52].

Based on the set of real performances, Fitts was able to apply a linear model in order to check how time to perform the task correlates with task difficulty. From **Figure 12**, we can see that a linear model can predict the time a tapping task will take in a given condition with a great degree of accuracy (R^2 = 0.966).

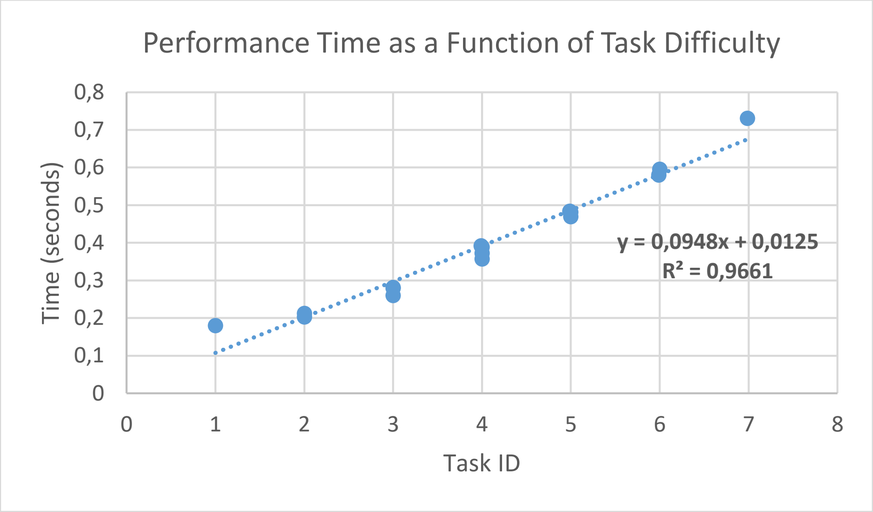


Figure 12. Linear Relation between the task ID and its time-to-completion. Data from [52], adapted using WebPlotDigitizer [55].

Thus, Fitts concluded that some perceptual-motor human behavior can be described by a liner function of the type:

Equation 2. $Movement\ Time\ (MT) = a + b(ID)$, *where MT is measured in milliseconds and where a and b are the intercept and slope, respectively, of a linear model fitted to a set of observed data.*

In order to draw the parallel between Fitts' Law and Information Theory equations, the above equation can be expanded as:

Equation 3. $MT = a + b\left[log_2\left(\frac{2A}{W}\right)\right]$, *where A is movement amplitude and W is target width.*

Fitts himself showed that his law could also be used to explain other perceptual-motor tasks involving grasp and releasing movements (such as disc transfer and pin transfer) [Fitts1954]. Given the robustness of its predictions and its focus on interaction performance, it was no wonder to see the first application of Fitts Law to HCI as soon as in 1978 [56]. Card, English and Burr compared four pointing devices – a mouse, a joystick, step keys, and text keys – for text selection in a CRT display (**Figure 13**).

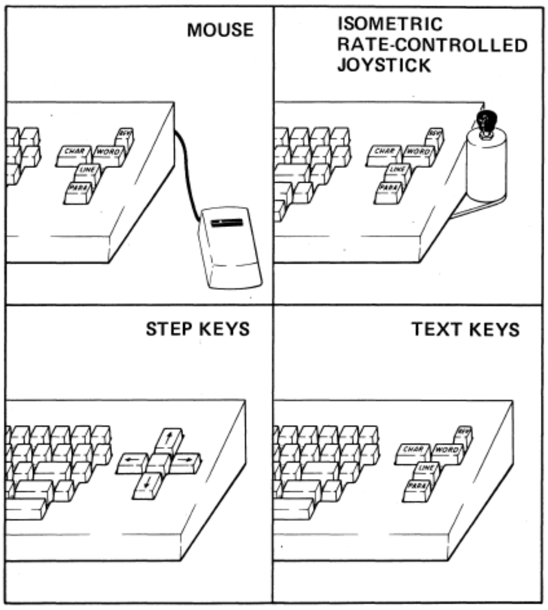


Figure 13. Pointing devices tested by Card and colleagues. Image from [56].

Similarly to the original Fitts' experiment, Card et al. [56] selection targets varied in distance to the cursor's initial position and in size (depending on the number of characters of the text target). The authors were able to compute different task's index of difficulty and manage to apply a linear regression to each one of the interaction conditions. The linear models accounted for more than 80% of the variation in time to reach the target as a function of the task index of difficulty, for all the interaction conditions. Thus, design choices based on PCMs were demonstrated feasible as the researchers could accurately predict how the user would perform with each one of the devices in different tasks of varying difficulty index.

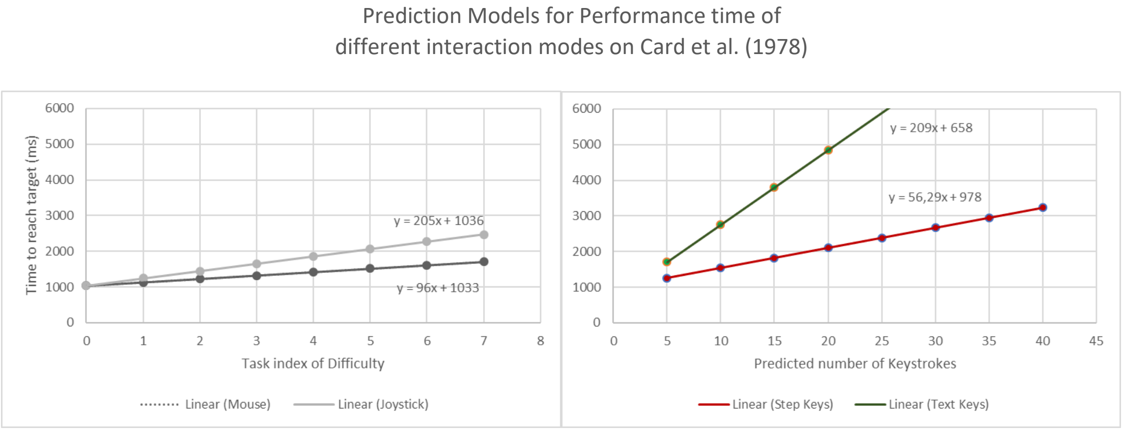


Figure 14. Linear Relation between the task ID and its time-to-completion, for a reaching task with different input devices – continuous input devices (right) discrete input devices (left). Data from [56], adapted using WebPlotDigitizer [55].

By applying Fitts' law models to their data, Card et al. were able to show that continuous input devices were a better design choice and that among those, the mouse allowed for the fastest and more accurate[17] level of performance (see **Figure 14**).

More recently Mackenzie and Jusoh [57] following the design choice perspective of Card and colleagues, tested performance on a typical reciprocal tapping task using a remote pointer *versus* a mouse. The distance to hover between targets were 40, 80, or 160 pixels, while the width of the target could be 10, 20, or 40 pixels. **Table 2** shows real performance and performance predicted by the Fitt's law models.

Table 2. Different task conditions in [57] and respective predicted and observed time to completion.

A	*W*	*W_e*	*ID_e*	*Real Time (ms) Mouse*	*Predicted Time (ms) Mouse*	*Real Time (ms) Remote Pointer*	*Predicted Time (ms) Remote Pointer*
40	10	11,23	2,322	665	605	1587	1473
40	20	19,46	1,585	501	487	1293	1222
40	40	40,2	1,000	361	362	1001	954
80	10	10,28	3,170	762	798	1874	1884
80	20	18,72	2,322	604	648	1442	1564
80	40	35,67	1,585	481	505	1175	1259
160	10	10,71	4,087	979	973	2353	2259
160	20	21,04	3,170	823	792	1788	1872
160	40	41,96	2,322	615	621	1480	1507
	Mean Time			***643***	***643***	***1555***	***1555***

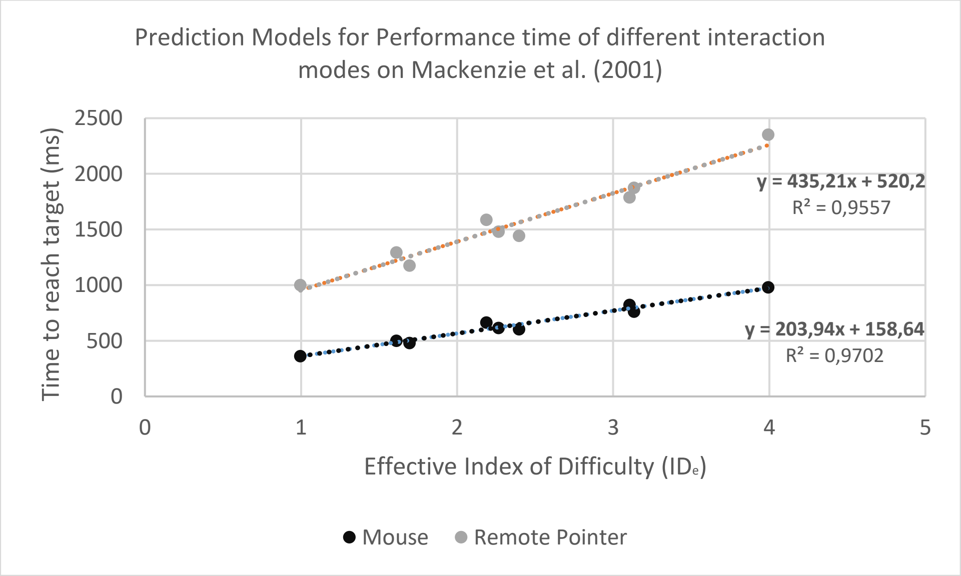


Figure 15. Linear Relation between the task effective ID and its time-to-completion, for a pointing task with different devices – Mouse (black dots) Remote pointer (grey dots). Data from [57], adapted using WebPlotDigitizer [55].

[17] Card, English, and Burr [56] also explore the error-rate for each type of interaction. Results rank devices in the same way as performance time.

Once again, it is interesting to point out the accuracy of the models in **Figure 15**, that yielded predictions that were on average precise to the millisecond (see last row on **Table 2**). This added accuracy might be due to a small, although important, change to the original Fitts' equation, proposed by Fitts in 1964 and extensively used by Mackenzie in different applications [58] [59] [60]. In 1964, Fitts suggested that when building a predictive model one should account for the variability of human motor performance during "hits" (or on-target) responses by defining the target width (*W)* as:

Equation 4. $W_e = 4.133 \times SD_x$, *where SD_x is the standard deviation of the distribution of "hits".*

Using what was named *effective width* (*W_e*) [53], we are considering motor variability by taking into account a consequent 4% probability for boundary errors and accommodating this variability by adapting the considered width of the target accordingly. This added improvement of the Fitt's models' predictions illustrates the continuously adapting nature of PCMs. By incrementally accounting for ever-detailed descriptions of human performance (based on incremental knowledge about human cognition and behavior), PCMs have the tendency to become more complete and to yield increasingly accurate predictions.

Hick-Hyman Law

PCMs are also suitable to explain intrinsically cognitive tasks as shown by the Hick-Hyman law for choice reaction time [61] [62]. Similarly to Fitt's Law, the Hick-Hyman law also takes the form of an information processing equation in order to make predictions about the reaction time in a task of target selection. In the original experiment Hick [61], presented the participants with a set of 10 pea lamps, arranged irregularly and connected to a set of 10 Morse keys (one dedicated to each finger). One random lamp would turn-on every 5 seconds and the participant had to press the associated key to turn it off. There was no logical arrangement between the lamp and associated key position and thus, participants had to first learn the association scheme. Crucial to the experiment, Hick presented the participants with several conditions that varied in the number of stimuli, ranging from sets of just 2 lamps to sets of 10 lamps (**Figure 16**).

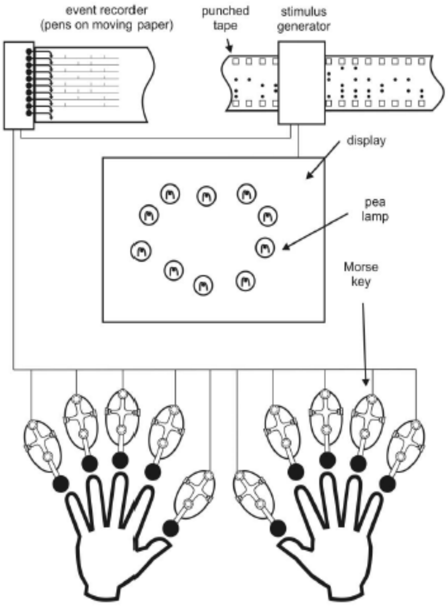


Figure 16. Author's rendition of the experimental apparatus based on Hick's 1952 experiment description, from [63].

Hick observed that average choice reaction time would increase logarithmically with the number of different options to choose from [61]. This relation could be predicted by the following equation:

Equation 5. $Choice\ Reaction\ Time\ (CRT) = a + b\ log_2(n+1)$, *where CRT is measured in milliseconds, a and b are constants empirically determined and n is the number of options.*

When plotting the real data of choice reaction time as a function of the degree of choice, Hick found that the data dispersion could be well explained by the equation $CRT = 0.518\ log(n+1)$. Thus, choice reaction time appears to be governed by an information theory principle, where additional alternatives function as added entropy and the human could be said to be functioning at almost full capacity in situations where the degree of choice was around 8 to 10 different options (**Figure 17**).

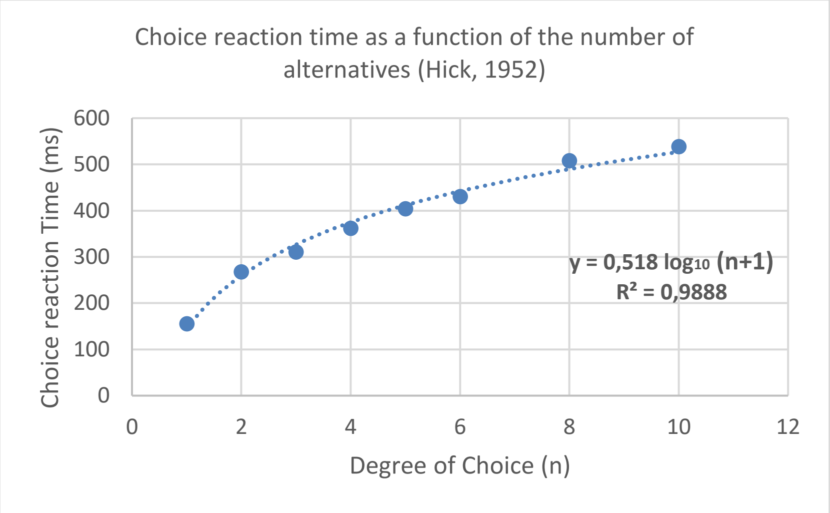


Figure 17. Data from subject A (Hick himself[18]) in experiment 1. Adapted from [61] using WebPlotDigitizer [55].

Almost at the same time Hick was conducting his experiments, Hyman was arriving to similar conclusions in a somewhat more complex experimental design [62]. Hyman's experiment had two parts: a first one where each lamp had the same probability to be turned on; a second part where probability of being turned on would differ from lamp to lamp allowing the participants to anticipate that certain lights had a higher probability of being turned on. This difference is quite important because it reflects the conditions of certain systems where specific operators, widgets, or telltales are more frequently involved in the interaction loop. For the information processing model describing choice reaction time, this distinction has some mathematical implications, namely on the part concerning system entropy: $log_2(n+1)$. Thus, for a set of alternatives with different probabilities, the information entropy value (H) should be:

Equation 6. $H = \sum_{i=1}^{n} p_i \, log_2\left(\frac{1}{p_i}+1\right)$, *where n is the number of alternatives and p_i is the probability of the i^{th} alternative.*

Consequently, if alternatives have equiprobable entropy, H is equal to: $log_2(n+1)$, and it is said the system works at maximum information rate or maximum entropy. Conversely, when alternatives are not equiprobable, the entropy of the stimuli is reduced [Seow2005] and choice reaction time improves, mainly because the user learns to anticipate certain occurrences. Thus, although Hick demonstrated that

[18] Although not acceptable by today's standards, it was common practice in the first half of XX century for experimental psychologists to participate as subjects in their own experiments. The rational was that by having devised controlled ways to measure cognitive, perceptual, or motor human performance, results would not be affected by knowledge about the experiment's goal. Iconic examples of self-experimentation were the pioneer of human memory study Hermann Ebbinghaus (1850-1909) and the social psychologist Stanley Milgram (1933-1984). For a reflection on self-experimentation see [236].

one could reduce entropy by reducing the number of alternatives, Hyman played with the probabilities of the stimuli to yield different amounts of entropy and study how choice reaction time could be described as function of *H* [63]. In its Experiment II, Hyman [62] comprised 8 different conditions, varying in stimuli set size and probability for each alternative, which resulted in the presentation of stimuli arrangements with entropy ranging from 0.47 to 2.75. **Figure 18**, shows how reaction time was affected by the different levels of entropy.

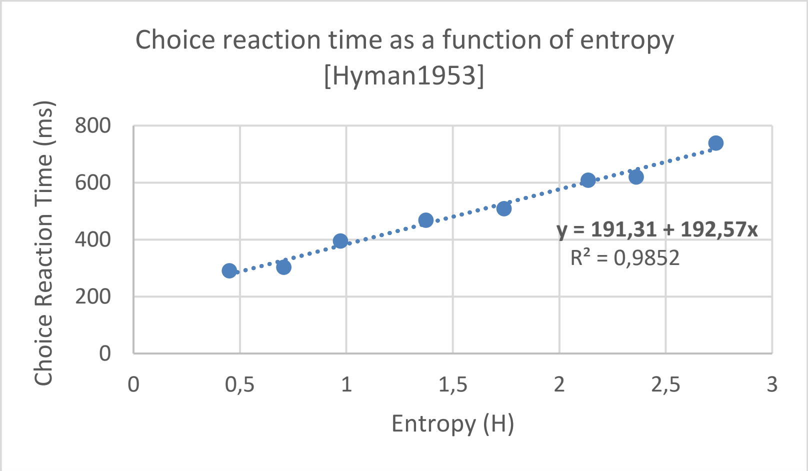


Figure 18. Data from experiment II. Adapted from [62] using WebPlotDigitizer [55].

Hyman results demonstrate once again how linear models can be used to explain and predict human performance, this time in a purely cognitive task (i.e., Hyman task did not require a motor action, instead participant's provided verbal answers). With the extension of Hyman's findings, Hick's law was renamed Hicks-Hyman law and is commonly represented in its linear form as:

Equation 7. $CRT = a + bH_t$, *where a and b are empirically determined constants and H_t is the entropy as defined in **Equation 6**.*

Implications of the Hick-Hyman law are obvious for information systems design. Knowing the performance ceiling in choice tasks can help developers and designers to adapt the interface by reducing the information content or by creating hierarchical presentations of the information, based on probability of use. Moreover, in safety critical systems where choice performance plays an important role (as in control rooms or in the operationalization of medical devices), warnings, alarm systems and mitigation procedures can be implemented taking into account the typical human choice reaction time. Considering the human cognitive capacity might help to prevent information overload in systems that require limited reaction time from the user.

PCMs based on the Hick-Hyman law were also found suitable to describe selection of items in hierarchical menus [64] [65], control displays [66] and data visualization [67], and to describe the operation of mode-based applications in a tablet interface [68]. Nevertheless, and as pointed out by Seow, while "Fitts' Law has enjoyed and continues to receive a great deal of attention in the field [HCI] the same cannot be said for the Hick-Hyman Law [that fail to] gain momentum in the field [63]". According to Seow, there are two main reasons why the Hick-Hyman Law has fallen short of the impact other models have caused in HCI:

1. *Difficulty in Application* – To apply the Hick-Hyman Law in its traditional formulation, one must first codify equivalent events into equiprobable or non-equiprobable alternatives in order to define information entropy. In today's complex interfaces, comprising a variety of multidimensional stimuli and different informational content, this is a difficult task to accomplish. This limitation often conveys the application of the law to simplified and highly controlled experimental set-ups which are unrepresentative of the current interfaces;

2. *Levels and type of performance* – While Fitts' Law captures and predicts a type of human performance mainly related to dexterity and motor behavior – which fostered its application in the study of control systems –, Hick-Hyman Law addresses a cognitive task with dimensions which are harder to directly observe and measure.

Despite their limitations, these influential PCMs made a lasting mark in HCI by first showing how controlled experimentations with user's yields valuable predictive models for both motor performance and cognition, allowing for informed design choices and incrementing the fundamental knowledge on interaction between humans and computing systems. In fact, Fitts' Law and Hick-Hyman Law became default principles in more ambitious human PCMs intended to describe the entire perception-action loop involved in human-computer interaction, such as the Keystroke-Level Model (KLM) [69].

The Keystroke-Level Model

In 1983, Card, Newell, and Moran, published one on the most seminal books in HCI called "*The Psychology of Human-Computer Interaction*" [70]. In their book, besides presenting the case for the articulation of cognitive and computer scientists in favor of the development of better user interfaces, Card et al. also presented a general PCM – the Keystroke-Level Model (KLM). Following the tradition initiated by Paul Fitts research, the KLM also uses the knowledge about the human motor system – inferred from simple experimental controlled tasks – as a basis for detailed predictions about user performance in compounded tasks. This model splits the execution phase into five different physical motor operators, a mental operator and a system response operator:

- **K** – Keystroking, the time it takes to strike a key;
- **B** – Pressing a mouse button;
- **P** – Pointing, moving the mouse (or a similar device) into a target;
- **H** – Homing, or the time that it takes to switch the hand between the mouse and the keyboard;
- **D** – Drawing lines using the mouse;
- **M** – Mentally preparing for a physical action;
- **R** – System response, which may be ignored if the user does not have to wait for it.

The execution of a task will involve interleaved occurrences of the various operators. The KLM predicts the total time for the execution of a task by adding the component times for each of the above activities, in such a way that the total time *T* is:

Equation 8. $T_{execute} = T_K + T_B + T_p + T_H + T_D + T_M + T_R$, *where the total time of each operator is the sum of all its occurrences in the described interaction process*

In many examples, the system response time is defined as zero – when needed, this response time can be measured by observing the system and be taken into account in the calculations. The times for the other operators all depend on the skills of the user and they can be obtained from empirical data with preliminary tests. Card, Moran, and Newell [69] gave us a general indication of the time for each operator, based on a sample of 1280 user-system-task interactions, comprised of various combinations of 28 users, 10 systems, and 14 tasks (see

Table 3). Although individual predictions may be interesting, the power of KLM lies in comparison between different systems. Having a detailed description of the methods to perform key tasks, we can use KLM to tell different systems apart by predicting, for example, which one allows for the faster interaction. This is considerably cheaper than conducting lengthy experiments and, furthermore, the systems do not even have to exist in a physical form.

Table 3. Times for operators in the KLM (adapted from [69]) and updated with [11] values for pointing.

Operator	Remarks	Time (ms)
K	PRESS KEY	
	Time varies with tipping skill:	
	Expert Typist (135wpm)	80
	Good typist (90 wpm)	120
	Average typist (40wpm)	280
	Typing complex codes	750
	Poor typist (unfamiliar with keyboard)	1200
B	MOUSE SCROLL	
	Down/Up	100
P	POINTING WITH A MOUSE	
	(Includes terminating mouse-button click)	$MT = 159 + 204\left[\log_2\left(\frac{A}{W} + 1\right)\right]$
	Fitts' law	
	Indicative average time	1100
H	HOMING KEYBORAD-OTHER DEVICES	400
D	DRAW n_DSTRAIGHT-LINE SEGMENTS OF TOTAL LENGTH l_D	$900n_D + 160l_D$
M	MENTALLY PREPARE	
	depending on the task type, Hick-Hyman Law might be applied	$CRT = a + b\left[\sum_{i=1}^{n} p_i\, log_2\left(\frac{1}{p_i} + 1\right)\right]$
	Indicative average time	1350
R	RESPONSE BY SYSTEM	
	System and command dependent values, measured empirically	t

In their original KLM experiment, Card and colleagues compared 14 tasks performed in different text and graphics editors. **Figure 19**, shows the predicted vs observed time of execution of each task. The position of the data points along the diagonal indicates that predicted and observed time are quite similar, staying under a 10% error rate in 12 out of 32 tasks.

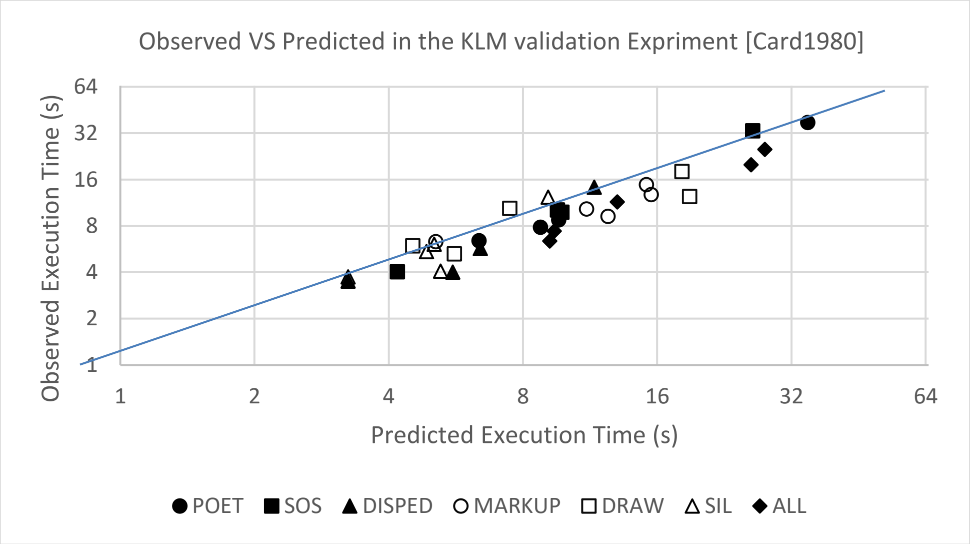
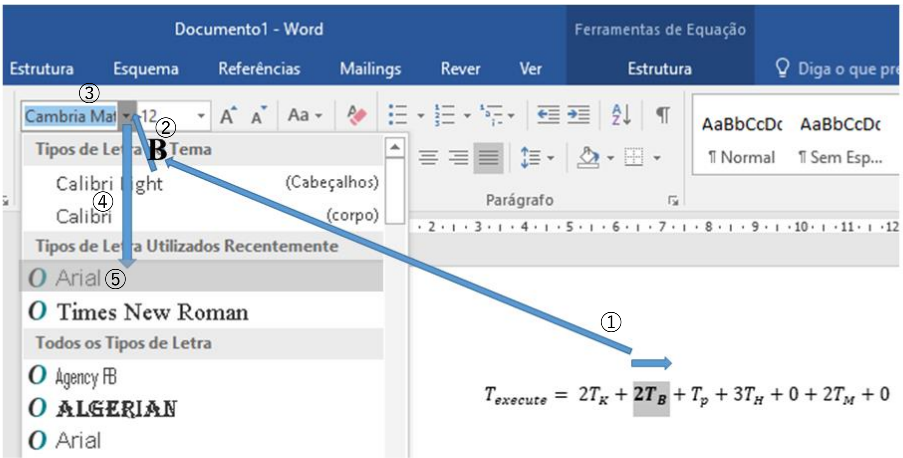


Figure 19. Observed versus KLM predicted time for text editing tasks. Data adapted from [69] using WebPlotDigitizer [55].

Using an example adapted from [11], we can understand the usefulness of KLMs for implementation comparisons and design choices. Consider the task of changing the second term on the equation depicted in **Figure 20** to boldface and Arial font, using (1) only the mouse and the Windows Word GUI, and (2) using only the keyboard.



$$T_{execute} = 2T_K + \mathbf{2T_B} + T_p + 3T_H + 0 + 2T_M + 0$$

Figure 20. Mouse operation for changing the second term of the equation to Bold and Arial font in Microsoft Word 2016®. ① to ⑤, represents the operation's order.

Changing to boldface and to Arial font using the mouse requires 4 pointing operators: selecting text; select Bold, select fonts drop-down, and select Arial font and 4 mental operators; while using only keyboard shortcuts requires 12 presses on key operators and 4 mental operators. Due to the high number of sub-tasks the user has to perform when using keyboard and the necessity to use combined keystroke commands, this task can be regarded as typing a complex code. If we compare the predicted time for the mouse performance and for the keyboard performance considering this a complex code task, we can expect significantly faster performances using the mouse (**Table 4**).

Table 4. Comparison of KLM predictions for the same editing task using mouse-based operations and keyboard-based operations.

Mouse Subtasks	**KLM Operators**	**Subtask time (s)**
Click and drag across text to select "$2T_b$"	M + P [2.5, 0.5]	1.35 + 0.686
Move pointer to Bold and click	M + P [13, 1]	1.35 + 0.936
Move pointer to Fonts drop-down and click	M + P [3.3, 1]	1.35 + 0.588
Move pointer down the list to Arial and click	M + P [2.2, 1]	1.35 + 0.501
	$\sum T_{Subtasks}$	**8.11s**
Keyboard Subtasks	KLM Operators	Subtask time (s)
Select Text	M + P [2.5, 0.5] + 3K[→]	1.35 + 0.686 + (3K)
Concert to Boldface	M + K[ctrl] + K[b]	1.35 + (2K)
Activate Format menu and Font sub-menu	M + K[alt] + K[o] + K[f]	1.35 + (3K)
Type "a" for Arial	M + K[a]	1.35 + K
Select Arial	K[↓] + K[Enter]	3K
	Typing complex codes setting $\sum T_{Subtasks}$	**14.4s**
	Average typist $\sum T_{Subtasks}$	**8.76**
	Good typist $\sum T_{Subtasks}$	**6.84s**
	Expert typist $\sum T_{Subtasks}$	**6.36s**

The KLM method has been empirically validated in a range of systems, both keyboard and mouse based, and in a wide variety of tasks. The predictions were found to be remarkably accurate [10] and thus, KLM stands as a quite interesting tool in the selection between different systems used in certain situations that require, for example, highly repetitive tasks or where the time necessary to perform a task is essential – such as telephony, data entry, and computer games.

One of the main drawbacks of the KLM is the difficulty to accurately define the correct time for the mental operator. Based on their observations, Card, Newell, and Moran suggested to use an average time of *1350 ms* (see

Table 3) or the use of Hick-Hyman Law, if the task is able to fit to the law assumptions (however this is task dependent). For novel tasks, this parameter has to be empirically measured in order to be added to the final PCM and, moreover, carefully consideration on where to put the mental operator on the task sequence has to be taken into account. Some software for calculating KML predictions (just as the *CogTool* [71] and *Cogulator*[19] [72]), automatically adds the values for the mental operator taking into account the particularities of the task. This, however, is based on expert behavior and seldom considers exploratory behavior, for instance. Teo and collaborators [73] adapted the *CoogTool* software in order to account for exploratory behavior of novice users, considering the type of UI. This adaptation resulted in predictions accounting for 63-82% of the variance in human performance.

Another limitation of the original KLM is the limited number of operators and their specialization on the desktop computer type of interaction. In order to expand the realms of application of the KLM, new work has been conducted on adding new interaction elements to the KLM. For example, [74] proposed new basic interaction elements for mobile phones and provided estimates for expert user performance derived from several user tests.

The Model Human Processor (MHP)

Fitts' Law, Hick-Hyman Law, and KLM are great examples of HCI models that illustrate the contribution that the study of human performance and human cognition can bring to the design of computerized systems. Nevertheless, being the firsts PCMs, they were quite limited in human cognition terms, often putting the emphasis in the human motor response and not doing significant assumptions about the human perceptual and processing mechanisms. However, in 1983 a leap forward was taken when again the team of Card, Moran, and Newell described the *Model Human Processor* (MHP), an ambitious attempt at simulating the multimodal human processing involved in interacting with computer systems. The novelty in this model is that it comprised three subsystems: the *perceptual system*, which handles information from the outside world; the *motor system*, which control actions; and the *cognitive system*, which provides the necessary processing to connect the two. The human information-processing is, in these models, conveniently described in terms of memories and processors, their parameters and interconnections. Thus, the model can be described as: (1) a set of memories and processors together with (2) a set of principles, called "principles of operation" [70], which are the operational analogies of the perceptual, cognitive, and motor mechanisms underlying human cognition and performance. Each

[19] Available at: http://cogulator.io/

one of these three subsystems has its own memories and processors and, in this sense, the perceptual system consists of sensors and associated buffer memories, the most important ones being visual and auditory image stores to hold the output of the sensory system while it is being symbolically coded. Then the cognitive system receives the coded information from the perceptual system in its working memory and uses information stored in the long-term memory to make decisions about how to respond. Finally, the motor system, after receiving a motor executive order carries out the adequate response.

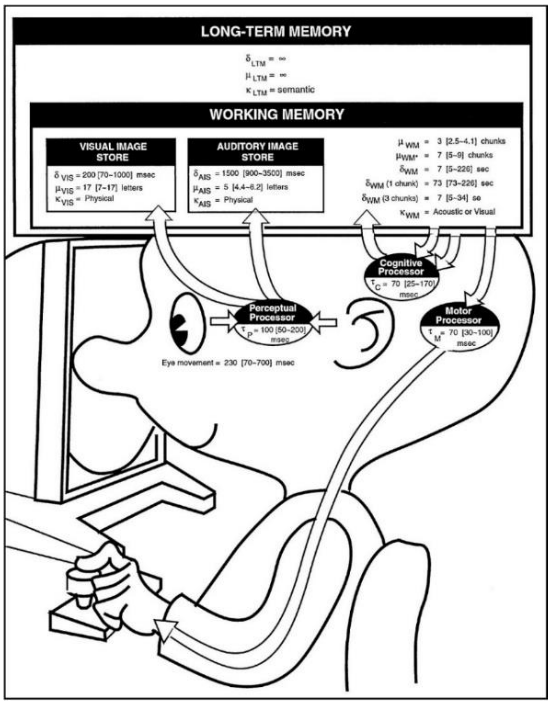

LONG-TERM MEMORY
$\delta_{LTM} = \infty$
$\mu_{LTM} = \infty$
κ_{LTM} = semantic
WORKING MEMORY
VISUAL IMAGE STORE
δ_{VIS} = 200 [70–1000] msec
μ_{VIS} = 17 [7–17] letters
κ_{VIS} = Physical
AUDITORY IMAGE STORE
δ_{AIS} = 1500 [900–3500] msec
μ_{AIS} = 5 [4.4–6.2] letters
κ_{AIS} = Physical
μ_{WM} = 3 [2.5–4.1] chunks
μ_{WM^*} = 7 [5–9] chunks
δ_{WM} = 7 [5–226] sec
δ_{WM} (1 chunk) = 73 [73–226] sec
δ_{WM} (3 chunks) = 7 [5–34] se
κ_{WM} = Acoustic or Visual
Cognitive Processor
τ_C = 70 [25–170] msec
Perceptual Processor
τ_P = 100 [50–200] msec
Motor Processor
τ_M = 70 [30–100] msec
Eye movement = 230 [70–700] msec

Figure 21. A depiction of the information process flow and some of the correspondent times, from [70].

The MHP conceptual basis uses a goal decomposition and hierarchy strategy. A high-level goal is decomposed into a sequence of unit tasks, all of which can be further decomposed down to the level of basic operators. This goal decomposition rational involves detailed understanding of the user's problem-solving strategies and of the application domain, a first input that can come from a more qualitative analysis typical of the DCMs (See **Section 4**). MHP predictions are based on a few parameters as memory storage capacity, memory decay constants and processing cycle's time that together with a set of well definer rules allow for the estimation of human performance times.

The beneficial application of MHP models to real-world design problems is quite clearly displayed in the paper by Gray, John, and Atwood, where they present an overview of Project Ernestine [75]. Gray and colleagues were called to perform a cognitive analysis in two systems of tele-operation after the New England phone company (now Verizon) realized that the proposed new workstation was costing an additional 4% in calls time. This was a quite counter-intuitive result since the new workstation was design with ergonomic and functional considerations that lacked in the design of the old workstation. Nevertheless, the MHP analysis conducted by Gray and colleagues predicted that the proposed new workstation would be 3% slower than the old one, mainly because the proposed new workstation placed more keystrokes on the critical path, rather than in the slack time, thus increasing the overall length of the call. In addition, the new grouping of the function keys, originally thought to reduce the distance to travel from key to key, encouraged the use of only the right hand to press all the keys, thus also contributing to a time increment in calls. The conclusion of their MHP analysis allowed the company to backtrack some decisions and clearly see were the problem was – and to save an estimated $2.4 million a year.

Despite their success, the biggest problem with the PCMs that are based on a rigid goal hierarchy structure is that these models were only fitted to describe how experts perform routine tasks. They are not of great practical use when we intend to describe more complex tasks which, for instance, can involve multiple-task performance. Moreover, a complementary problem is that the MHP and similar PCMs rely on robust psychological results but ignore second-order phenomena usually responsible for specific effects on human perception and performance (e.g. differentiation of the reaction time to visual stimuli depending if the stimulus is presented to the central vision or the peripheral vision). However, when one looks at human cognition one enters an all-new level of complexity, and in order to give accurate predictions of human performance in a multisensory task, a PCM has to model, accurately, the intricacies of human multimodal perception and action. Therefore, and in order to provide more insight on more recent PCMs, we will briefly describe the *Executive Process-Interactive Control* (EPIC) [76] and the *Adaptive Control of Thought-Rational* (ACT-R) [77], cognitive architectures that were designed to account for some of the MHP debilities.

The EPIC & ACT-R

Despite being conceptually similar to MHP, both the EPIC and the ACT-R architectures updates many theoretical and empirical results about human perception and performance and, consequently, are better fitted to the simulation of both low-level and complex interaction tasks in multimodal domains. There are two fundamental differences between MHP and the more recent cognitive architectures, that render the latter as more successful attempts at representing the perceptual, motor, and cognitive constrains on human ability to interact with computerized systems:

1 EPIC and ACT-R takes seriously the notion of "*Embodied Cognition*". In the previous PCMs, the human is represented as a "...*purely cognitive system, a disembodied intelligence that directly perceives and acts on its environment*." [76] (p. 392). Such an approach neglects the human body constraints. Consider, for instance, the human capacity to detect and recognize objects. This capacity is not uniform: we do not take the same amount of time to recognize each and every visual stimulus. Instead this time varies with the distance on the retina from the fovea (peripheral stimuli takes longer to be coded) [76] [50]. Moreover, it can vary with the level of familiarity that the user has with an object, or a word, and it also can vary depending on previous interactions [78]. These kind of details, have seldom been considered in the development of predictive human cognitive models and the fact is that they can have a huge impact in the prediction output. The embodied cognition architectures take these considerations a step further than precedent models. For instance, it goes beyond MHP by specifying separate perceptual processors with distinct processing times for each sensory modality, and separate motor processors for vocal, manual, and oculomotor movements. There are also feedback pathways from the motor processors, as well as tactile feedback from the effectors, which are important in coordinating multiple tasks [76];

2 The second fundamental difference is that these embodied cognition architectures, can have several active rules (generated by the production-rule system in the cognitive processor) at one time in the working memory. This makes it possible, for example, to have the cognitive processor activating some manual response in the effectors while sending orders to the perceptual processors in order to change the content of the visual or auditory working memory. Thus, these architectures allows for the description of cognitive processes that run in parallel and can account for multitasking activities.

EPIC has a production-rule cognitive processor surrounded by perceptual-motor peripherals processors (see **Figure 22**). There is a conventional flow of information from sense organs, through perceptual processors, to a cognitive processor, and finally to motor processors that control effector organs. The perceptual processors are simple "pipelines", in the sense that an input produces an output to working memory within certain latency time. In this way, it is possible to simulate the natural processing delays occurring in humans' perceptual systems, which have to take into account the different transduction times for different receptor organs and possible different processing time dependent on the cognitive task. In addition to having different delays for different sensorial modalities, EPIC also has different delays for different kinds of visual events – depending if they are *foveal*, *parafoveal*, or *peripheric* – and for the processing of different characteristics of a sound. The cognitive processor, in its turn, operates cyclically, with a mean duration of 50ms per cycle, and it is programmed in terms of production rules. Thus, an EPIC model for a task must include a set of production rules that specify what actions in what situations ought to be performed. The production rules are a sequence of **if** *condition* **then** *action* rules written in Common LISP (see **Example 1**) and the rules' conditions can only test the contents of the production system working memory. Action rules can add or remove items from the working memory, and send a command to a motor or a perceptual processor. Finally, the motor processors can produce a variety of simulated movements of different effector organs, taking varying amounts of time to do so. The movement features remain in the motor processors' memory, so that future movements that share the same features can be performed more rapidly. Although a motor processor can only prepare the features for one movement at a time, the preparation for a new movement can be done in parallel with the physical execution of a previously commanded movement.

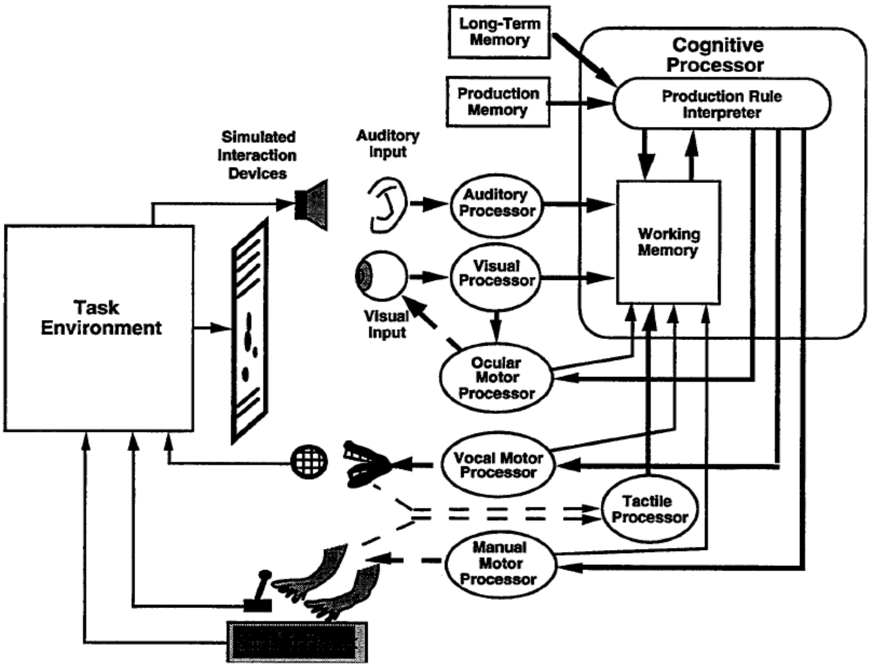
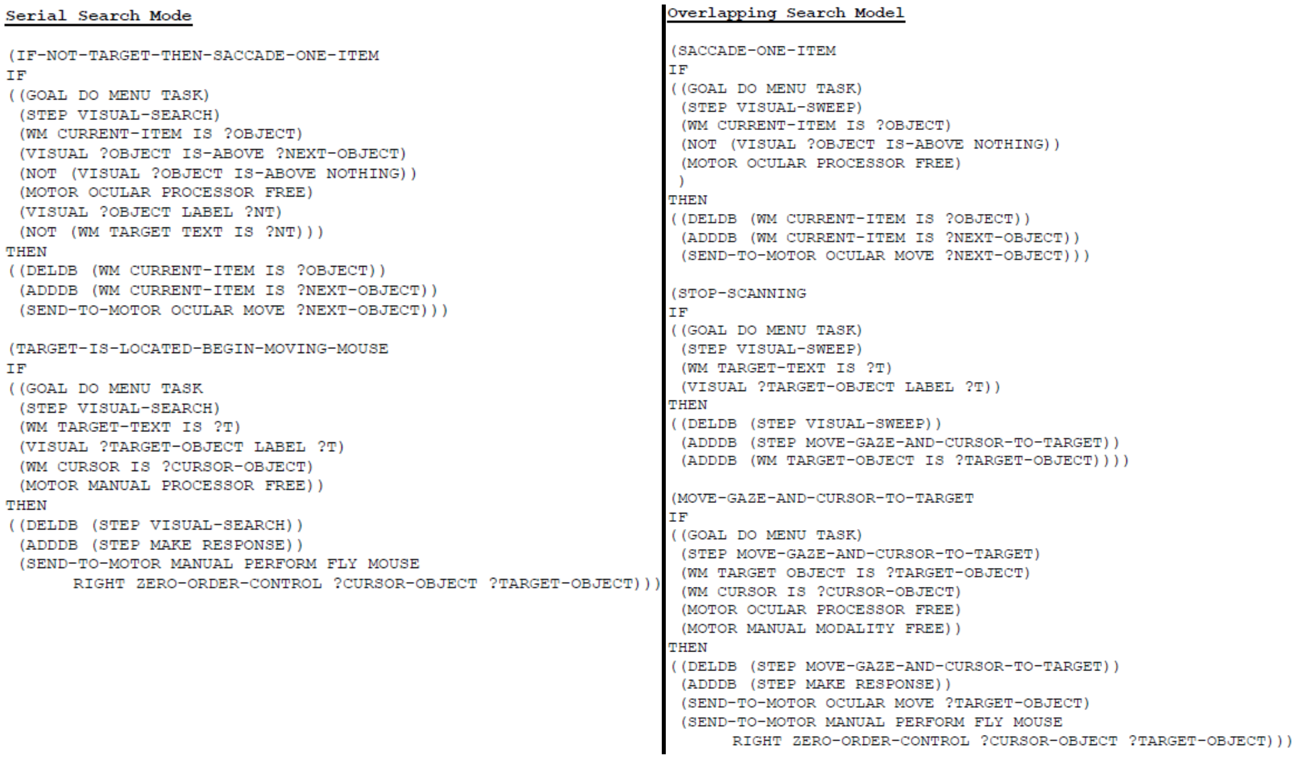


Figure 22. Overall structure of the EPIC architecture simulation system. Information flow paths are shown as solid lines, and mechanical control connections as dashed lines [76].

An illustrative application of the EPIC architecture is presented in [76]. They thought about a typical activity in desktop environments – selecting items from menus – and modeled performance in a task that was one condition of an experiment by Nilsen [79]. In Nilsen's experiment, a digit was first showed to the subject who would then proceed by clicking on a target, causing a vertical menu of digits (from 1 to 9) to appear in a random order below the cursor. The subject's experimental task was to point and click on the previously showed digit. Nilsen's results showed that the time to select the correct digit was a function of its location on the randomly ordered menu, and that it was a fairly linear function with a slope of about 100ms per item. Modeling Nilsen's task in EPIC required estimating a parameter for how long it takes to recognize the text label for digits, and where on the retina this recognition happens. Then, two models were constructed: (1) a *serial search model* that corresponded to a one-at-time hypothesis checking for visual search, where the eye is moved to the next object down the menu and if it matches the sought for target-item, a pointing movement to the that item is initiated; (2) an *overlapping search model* where the parallel processing possible in EPIC is fully exploited because two functions are actively working at the same time, one that is moving the eye as rapidly as possible from one item to another (SACCADE-ONE-ITEM) and the other that is monitoring the work memory (STOP-SCANNING, see **Example 1**) expecting the emergence of the target item to stop the search and start the movement of the cursor to the target.



```
Serial Search Mode

(IF-NOT-TARGET-THEN-SACCADE-ONE-ITEM
IF
((GOAL DO MENU TASK)
 (STEP VISUAL-SEARCH)
 (WM CURRENT-ITEM IS ?OBJECT)
 (VISUAL ?OBJECT IS-ABOVE ?NEXT-OBJECT)
 (NOT (VISUAL ?OBJECT IS-ABOVE NOTHING))
 (MOTOR OCULAR PROCESSOR FREE)
 (VISUAL ?OBJECT LABEL ?NT)
 (NOT (WM TARGET TEXT IS ?NT)))
THEN
((DELDB (WM CURRENT-ITEM IS ?OBJECT))
 (ADDDB (WM CURRENT-ITEM IS ?NEXT-OBJECT))
 (SEND-TO-MOTOR OCULAR MOVE ?NEXT-OBJECT)))

(TARGET-IS-LOCATED-BEGIN-MOVING-MOUSE
IF
((GOAL DO MENU TASK
 (STEP VISUAL-SEARCH)
 (WM TARGET-TEXT IS ?T)
 (VISUAL ?TARGET-OBJECT LABEL ?T)
 (WM CURSOR IS ?CURSOR-OBJECT)
 (MOTOR MANUAL PROCESSOR FREE))
THEN
((DELDB (STEP VISUAL-SEARCH))
 (ADDDB (STEP MAKE RESPONSE))
 (SEND-TO-MOTOR MANUAL PERFORM FLY MOUSE
      RIGHT ZERO-ORDER-CONTROL ?CURSOR-OBJECT ?TARGET-OBJECT)))
```

```
Overlapping Search Model

(SACCADE-ONE-ITEM
IF
((GOAL DO MENU TASK)
 (STEP VISUAL-SWEEP)
 (WM CURRENT-ITEM IS ?OBJECT)
 (NOT (VISUAL ?OBJECT IS-ABOVE NOTHING))
 (MOTOR OCULAR PROCESSOR FREE)
 )
THEN
((DELDB (WM CURRENT-ITEM IS ?OBJECT))
 (ADDDB (WM CURRENT-ITEM IS ?NEXT-OBJECT))
 (SEND-TO-MOTOR OCULAR MOVE ?NEXT-OBJECT)))

(STOP-SCANNING
IF
((GOAL DO MENU TASK)
 (STEP VISUAL-SWEEP)
 (WM TARGET-TEXT IS ?T)
 (VISUAL ?TARGET-OBJECT LABEL ?T))
THEN
((DELDB (STEP VISUAL-SWEEP))
 (ADDDB (STEP MOVE-GAZE-AND-CURSOR-TO-TARGET))
 (ADDDB (WM TARGET-OBJECT IS ?TARGET-OBJECT))))

(MOVE-GAZE-AND-CURSOR-TO-TARGET
IF
((GOAL DO MENU TASK)
 (STEP MOVE-GAZE-AND-CURSOR-TO-TARGET)
 (WM TARGET OBJECT IS ?TARGET-OBJECT)
 (WM CURSOR IS ?CURSOR-OBJECT)
 (MOTOR OCULAR PROCESSOR FREE)
 (MOTOR MANUAL MODALITY FREE))
THEN
((DELDB (STEP MOVE-GAZE-AND-CURSOR-TO-TARGET))
 (ADDDB (STEP MAKE RESPONSE))
 (SEND-TO-MOTOR OCULAR MOVE ?TARGET-OBJECT)
 (SEND-TO-MOTOR MANUAL PERFORM FLY MOUSE
      RIGHT ZERO-ORDER-CONTROL ?CURSOR-OBJECT ?TARGET-OBJECT)))
```

Example 1. Production rules for the serial search model (left) and for the overlapping search model (right), from [76].

Figure 23 shows the results from the predictions made by the two models using EPIC, and the results from Nilsen's experiment with humans. As we can see, there is a considerable difference between the *serial-search* and the *overlapping-search model* for the item selection time predicted as a function of the item position. Clearly, the model that best explains the real human performance is the *overlapping-search model*, a model that exploits in a more satisfactory way the multitasking nature of some human performances. Once again, this result reveals the ever-improving nature of PCMs, which tend to improve their predictions with the increment of the accurate and completeness of their models of human cognition.

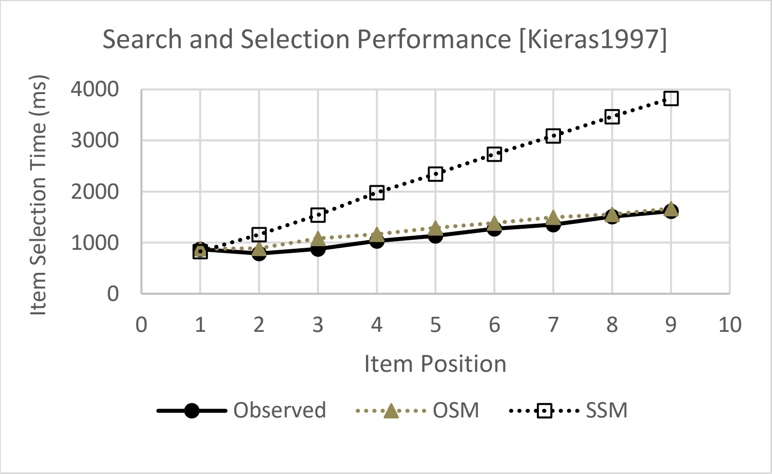


Figure 23. Menu selection times of the observed data from [79] with the predicted times by two EPIC models [76], the Serial Search Model (SSM) and the Overlapping Search Model (OSM).

Similar to EPIC, the ACT-R also consists of a set of modules dedicated to a specific kind of information but capable of contributing with outputs for the integrated cognition process. [77] illustrates the basic architecture of ACT-R as a *visual module* – capable of identifying objects in the visual field –, a *manual module* – to issue commands and control the actuation of the hands –, a *declarative module* – for retrieving information from the memory – and, finally a *goal module* – to keep track of the current goals and long-term intentions (**Figure 24**).

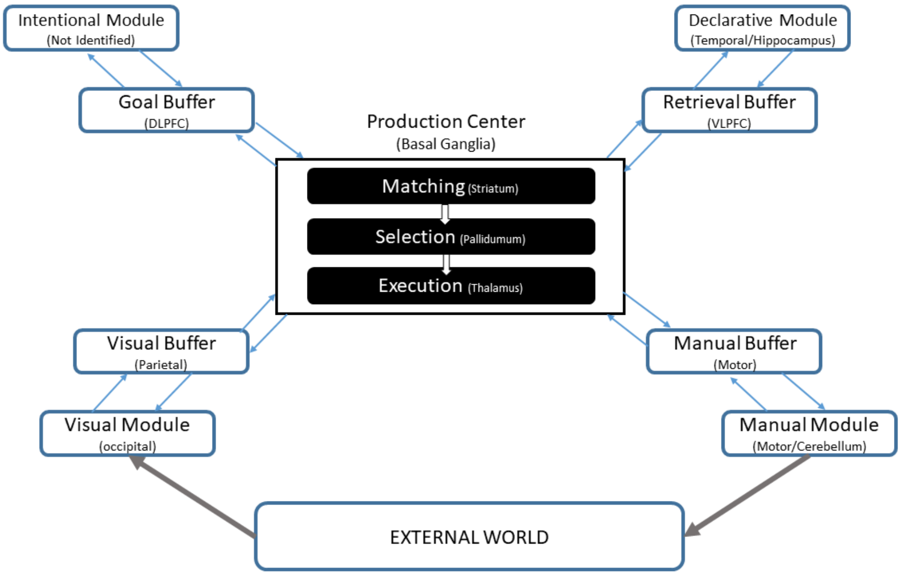


Figure 24. Information architecture of the ACT-R version 5.0 adapted from [77]. In parentheses are main brain regions where these functions have been identified and that the ACT-R attempts to modulate.

The coordination of these modules is achieved by a set of rules executed in the *Central Production System* (CPS), intended to modulate the functioning of our central nervous system. The CPS is not sensitive to most of the activity occurring in the different modules, but rather it only responds to the information deposited in the buffers of those modules. The original architecture of the ACT-R was not committed with how many modules existed and consequently new modules or production mechanisms were implemented as part of the core system. For instance, in 2006 a new version of the ACT-R code was released (ACT-R 6.0) which included a new mechanism in the CPS called dynamic pattern matching [80], and more recently dedicated modules for speech and audition where also developed. Similarly to EPIC, the ACT-R is available in a Common LISP distribution[20], and its framework allows developers to create new modules or partially modifying the behavior of the existing ones, by changing their standard parameters.

The general accuracy of some PCMs is, in fact, notable and it is highlighted here because it demonstrates quite well the mutual contribution between cognitive and computer sciences. PCMs in HCI tried to integrate aspects of perception, attention, short-term memory operations, planning, and motor behavior in a single (executable) model, at a time when most cognitive science models addressed only

[20] ACT-R software can be downloaded here: http://act-r.psy.cmu.edu/software/; an implementation in Python can be downloaded here: https://github.com/jakdot/pyactr

isolated laboratory phenomena [7] – and this attempt obtained some success. In an almost symbiotic way, HCI PCMs have also been benefiting immensely from periodic actualizations based on the latest advances in experimental psychology. As we saw, the tuning of these interaction models often relies on the definition of fixed and variable parameters that are only accessible through psychological and human factors experimentation on real users. This relationship of mutual benefit is becoming increasingly clear for professionals of both areas, and the future of both areas may lay on the fact that they depend on each other – technology designed for humans will always require human studies and the knowledge frontier on human cognition and performance may only be extended using new technology.

Having in mind this symbiotic relation between cognitive and computer sciences within the HCI field, we can say that one of the most important questions for the work presented in this thesis is, *what can be the practical use of merging experimental psychology and computer sciences on the development of immersive environments?*

As we could see from the PCMs presented before, the knowledge about human characteristics can add a lot to the modeling of human performance and subsequent design and implementation in computerized systems that require human interaction. By looking at the current state of IVEs, we will get a clear sense of the existing limitations and of how our approach can help to surpass them. The importance of adopting an iterative logic of user evaluation / re-design (described in more detail in **Figure 4**), is even more clearly understood when we analyze the limitations that current IVEs still present. By bringing back the empirical approach of the first HCI researchers of the like of Fitts, Card, and Mackenzie, and by resorting to controlled user testing, we will propose new PCM and design guidelines targeted at improving such IVEs. At a time when IVEs are being considered as the next revolution in terms of UI, it is only natural that the HCI principles of controlled user testing are transposed, adapted, and applied to IVEs development and research.

2.6. The current state of IVEs technology

Visual IVEs

As pointed out by Jerald [24], today's visual IVEs are implemented in one of three ways: HMDs, world-fixed displays (screen or projection-based displays), and hand-held displays. In this section, we will review the first two, since hand-held displays (such as tablets) are out of the scope of highly immersive environments.

HMDs are, indisputably, the currently most explored support in the development of IVEs. In 1984, the NASA Ames research center presented the *Virtual Visual Environment Display* (VIVED), a monochromatic stereoscopic HMD, capable of displaying computer-generated images (CGI) with a frame rate that varied from 20 Hz up to 60 Hz depending on the complexity of the graphics [81]. The use of HMDs as the preferred support for IVEs continues today, as Oculus Rift® became the first enterprise to successfully commercialize this type of solution as a consumer good (**Figure 25**).

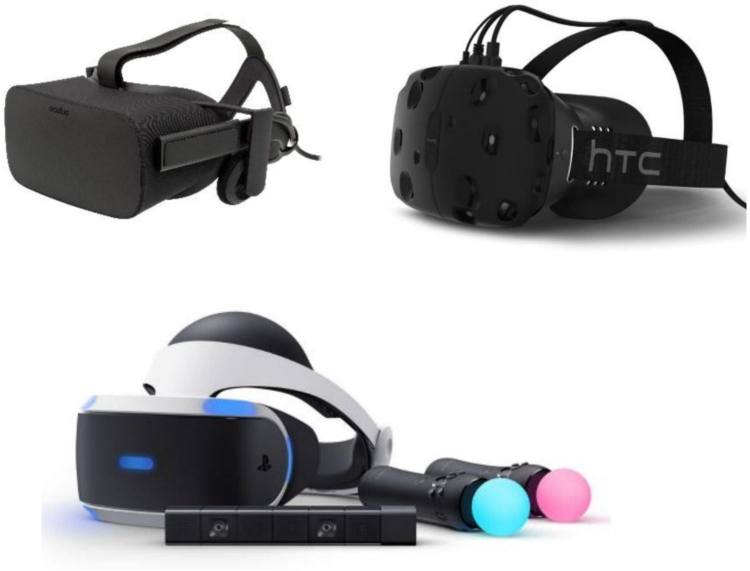


Figure 25. PC-based premium HMDs currently commercially available. Top-Left is the Oculus Rift Consumer Version 1; top-right is the HTC Vive; down is the PlayStation VR from Sony (images distributed under a CC-BY 2.0 license).

Independently of the type of displays, an HMD requires the presentation of different images to each eye, thus making extensive use of binocular cues (binocular disparity and convergence) to provide visual depth perception. HMD's are comparably easy to transport and set-up and its price is also an important advantage when compared with other projection-based or world-fixed displays systems. Moreover, HMDs have been developed to address all the spectrum of the R-V Continuum with purely VR, mixed reality, and AR devices already on the market (**Figure 25**).

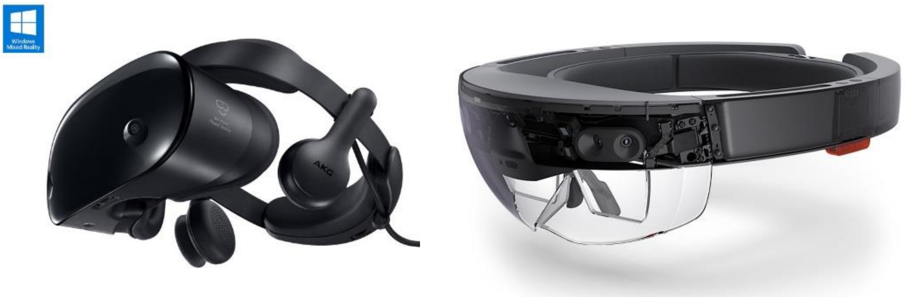


Figure 25. The Samsung Odyssey and the Microsoft HoloLens, two of the state-of-the-art equipment for mixed reality and augmented reality.

However, the problems and limitations of HMD are many. We have already referred some of them in **Section 2.2**, nevertheless, we will provide next some additional detail and link the limitations to different types of HMDs technology. One general limitation is the end-to-end delay, with current performance failing to achieve the 15-20ms value that might prevent motion sickness [31]. Nevertheless, threshold values for latency perception vary greatly with the experimental condition (i.e. type of head movement, differences in the experimental protocol, and individual variability) [82]. Having built a laboratory HMD with an end-to-end latency of 7.4ms, Jerald [83] identified a participant capable of noticing a delay of just 3.2ms. In AR HMDs, where the real and synthetic images are merged in space and time, the latency problem becomes even more visually noticeable because any latency will result in displacement between CGI and real world scenarios. [24], suggest that the threshold should amount to no more than just 1ms for this type of displays, mainly due to the possibility of participants judging latency in relation to real-world fixed objects. A list of negative effects of latency is provided by [24] and, in addition to motion sickness, also includes: degraded visual acuity; degraded performance; breaks-in-presence; and negative training effects.

A second problem with optical see-trough HMDs is light intensity. Light conditions of the real and the virtualized world have to be carefully controlled; otherwise, if the synthetic imagery is too bright relative to the ambient light, the real environment will not be visible – the reverse being also true. A final difficulty with optical see-trough displays is that of occlusions. If a virtual object should appear in front of a real-world object, it will generally appear to be semitransparent (holographic-like) and this transparency increases as a function of the occluded real-world object's luminosity.

Fuchs and Ackerman [84], described an interesting problem of visual displacement in mixed reality HMDs, an equipment that conveys to the user a view of the real world through one or more video cameras mounted on the front of the HMD. A mixed-reality HMD merges virtual imagery with camera recordings of the real world, thus giving the user a completely CGI environment. The problem is that without carefully optical considerations, it is quite difficult to align the camera's view with the normal viewing axis of the user's eye. The image sent to each of the user's eyes is taken from a camera perspective other than that of the observer's eyes, and this could distort the user's sense of depth because the stereo pairs have an effective pupillary distance different from that to which the user is accustomed. The interesting perceptual result of this incongruence is that the user might get a false sense of height, particularly for near field objects.

Other performance limitations are general to all current HMDs, starting with the *Field of View* (FoV) problem. Humans have a horizontal FoV of about 160 degrees (≈114 degrees of binocular vision, ≈40 degrees of peripheral non-binocular vision) and a vertical FoV of about 120 degrees [50]. The FoV has been linked to a greater sense of immersion ([85], [86]), nevertheless current HMDs are still limited regarding this aspect. Information available about the FoV specifications of existing devices is contradictory, with specifications from manufacturers differing from measurements performed by independent users and technology reviewers. One of the few reviews that presents values for both horizontal and vertical FoV indicates a horizontal value of 100° FoV and a vertical value of 110° FoV for the HTC Vive, while the Oculus Rift CV1 presents values around 80° FoV horizontal and a vertical value of 90° FoV[21]. Published controlled measurements are scarce, nevertheless, [86] found values around 90° FoV for the Oculus Rift CV1. For the AR headset Hololens, the FoV limitation is even more problematic with a reported horizontal value of 30° FoV and a vertical value of 17.5° FoV.

The Headset-fit and weight are also reported as important usability issues affecting HMDs [24]. Loose-fitting HMDs can cause unwanted small scene motions, while overly tight HMD will cause too much head pressure and discomfort. This problem is positively correlated with the weight of the equipment, which is around 470 grams for the Oculus Rift CV1 and the HTC Vive and can amount to around 579 grams for the Microsoft HoloLens.

IVEs based in world-fixed displays, have the capacity to deal with the generality of the usability issues linked to HMDs, since they remove the physical affixation of the projection to the user's head, nevertheless they are not immune to the perceptual problems that also affect HMDs. In 1993, Cruz-Neira, Sandin, and DeFanti presented their work on the design and implementation of the Cave Automatic Virtual Environment (CAVE), and according to these authors, the CAVE came (at the time) as the first system that fully met the standards that defined a VR system. In their words:

> "... *a VR system is one which provides a real-time viewer-centered head-tracking perspective with a large angle of view, interactive control, and binocular display.*" [26] (p.135).

Contrary to most of the prior technology, the first CAVE presented at SIGGRAPH '93 was conceived as a tool for scientific visualization and, therefore, high performance and accuracy in the representation of the real world was needed in order to convince leading-edge computational scientists to use the CAVE

[21] Review accessible on https://www.vrheads.com/field-view-faceoff-rift-vs-vive-vs-gear-vr-vs-psvr

in alternative to other more traditional immersive environments. In this sense, the goals that inspired the CAVE engineering effort were:

1. The desire for higher-resolution color images and good surround vision without geometric distortion;
2. Less sensitivity to head-rotation induced errors;
3. The ability to mix VR imagery with real devices (like the user hand, for instance – which leaves out the necessity of *DataGloves*-like equipment or other peripherals to interact with the virtual environment);
4. Minimize the existence of attachments and disturbing hardware around the user;
5. The desire to couple to networked supercomputers and data source for successive refinement.

The original CAVE, developed at the University of Illinois was a 3x3x3 meters cubic structure, comprising three projection screens on vertical walls (the front one and the two lateral ones, with respect to the user) and one projection screen for the floor.To "transform" cube sides in projection planes, a CAVE system frequently uses a *Window Projection Paradigm* in which the projection plane and projection point relative to the plane are specified, thus creating an off-axis perspective projection. To get a differentiate image to each eye separately, the most common solution is to use frame sequential stereo with synchronized shutter glasses, a solution that immensely reduces the flicker effect. Images are rear projected in the walls, so that participants in the CAVE do not cast shadows on the projection screens – only in the ground, when the projection comes from the top of the user. The user's tracking was originally performed with Polhemus™ tethered electromagnetic sensors, but currently it is made resorting to high precision motion tracking optical systems, such as the Vicon™ or the Qualisys™ systems. By combining high precision tracking with high performance computation[22], CAVE systems are ideal to convey IVEs with high-resolution images and lower end-to-end latencies.

CAVE systems might constitute a solution for the limited FoV problem and for the usability issues of obstructive hardware on HMDs. A CAVE system provides a surround projection and requires no peripherals other than lightweight stereoscopic glasses and retroreflective markers for user tracking. Nevertheless, a CAVE configuration is not exempted of usability issues. One of the most relevant issue is the space the user is confined to, often a 3m^2 surface surround by 3m^2 walls. Thus, natural navigation

[22] In world-fixed displays IVEs computers can be housed further from the display, thus avoiding any space limitation. State of the art graphic boards and processors can be used to construct computer clusters that might optimize the performance of the IVE assuring high resolution images and low end-to-end latency.

(i.e., walking) in the real room is restricted and the user has to resort to commands in order to navigate the virtual space. Some CAVEs have dwelt with this problem by applying omnidirectional treadmills (**Figure 26**), thus allowing for continuous walking. However, this solution also carries some problems. As we can see in **Figure 26**, the user as to use a safety harness, adding obstructive equipment to the simulation, and the treadmill control to adapt to some natural walking patterns such as accelerations, decelerations and turns is quite difficult to obtain mainly due to the necessity of predicting user's waking pattern to adapt timely [87].

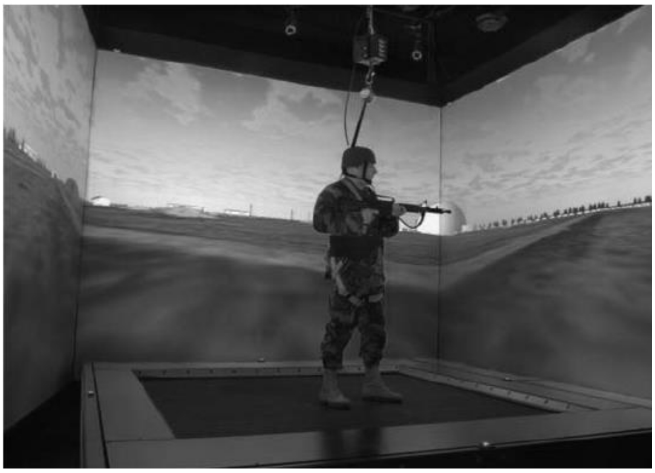


Figure 26. A CAVE visualization system with an omnidirectional treadmill from the U.S. Army Research Lab. From [88]

A more convenient solution for the space restriction problem is the adoption of a *Power-Wall* configuration. By simply opening the left and right sides of a CAVE, one can get a power-wall of about 9 meters wide by 3 meters height (**Figure 27**). This allows for plenty of additional room to lateral movements (though maintaining the same limitation in forward movements), while keeping FoV values considerably high (FoV values will increase as the user's get closer to the projection surface).

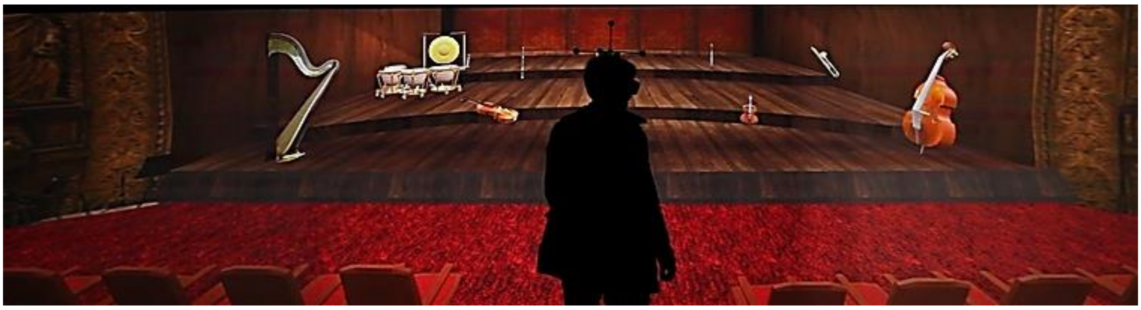


Figure 27. The author using a CAVE-like system in a power-wall configuration. Image from the Visualization System located at the Center for Computer Graphics (CCG), Guimarães-Portugal.

Recently a new configuration was developed for CAVE systems, which is an interesting combination of the traditional CAVE and the power-wall solution. In 2018 Antycip Simulation, a French company that develops visualization systems, presented the TORE (The Open Reality Experience) (**Figure 28**). This system was first installed in Lille University and is described as an edge-less CAVE capable of completely surrounding the user, while still providing considerable space for natural movement.

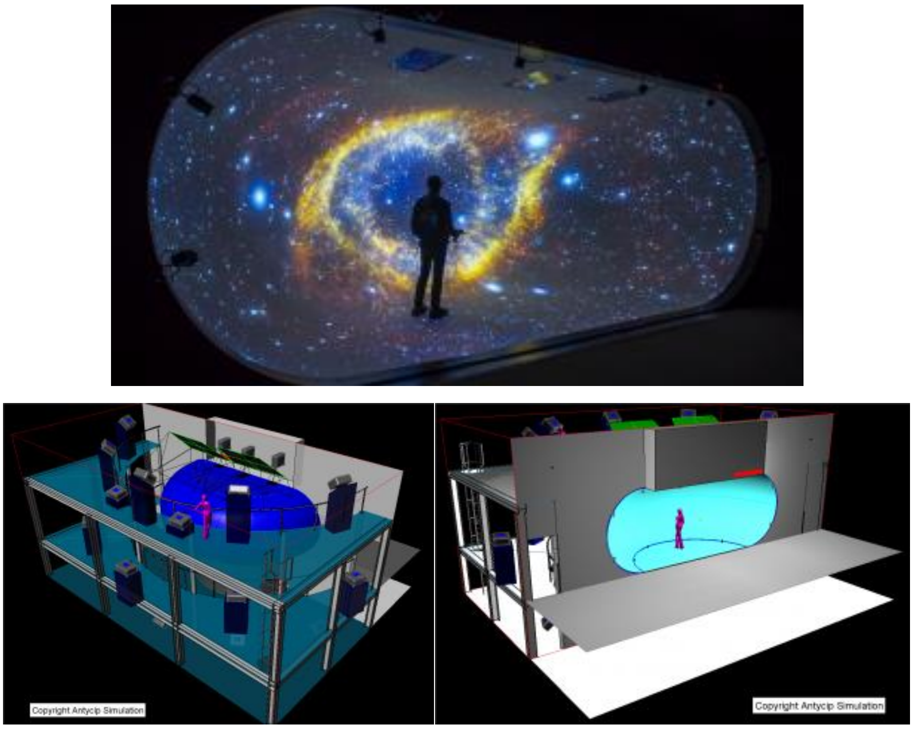


Figure 28. The TORE system at Lille University, France. Images from Antycip Simulation.

Although impressive and clearly beneficial in usability terms, CAVE-like systems have the main drawback of being highly expensive to acquire and to maintain. A CAVE or a power-wall that requires three projectors can cost around 150 000 € for the projection system alone, and the TORE system installed in Lille University required 20 high-resolution 3D projectors. Thus, contrary to what is starting to happen with HMDs, CAVE visualization systems are not on the spectrum of a consumer good, but are almost exclusively used for research applications, military training, driving simulators, and as part of permanent exhibitions on museums, planetariums, and scientific/educational centers.

After revisiting some of the most important visual IVEs technology that were developed since Sutherland's "ultimate display" we can safely state that the Sutherland challenge of a true VR environment has not been yet accomplished. Although, a CAVE-like environment is the best candidate for the visual component of a full VR environment since it deals with most of the usability problems of HMDs, this type of system still constitute quite expensive solutions, and are not immune to some perceptual problems mainly arising from the link between user's tracking and image projection and the problem of end-to-end delays. The CAVE developers warned for the necessity of design and implementation of quantitative experiments to measure CAVE performance [26], in other words, the need for proper HCI evaluation in order to successively refine the CAVE design. In this sense, psychophysical studies of visual depth perception and of the human temporal recalibration mechanisms (mechanisms that allow us to deal, for example, with inter-modal temporal lags) should guide the development of PCMs that allow for improvements in CAVE-like systems implementation.

Acoustic IVEs

One main problem of current IVEs is the one-modality prevalence of the majority of the commercially available systems. [26] have previously identified the necessity of visual and aural integration in VR environments but, despite this fact, few advances were made in that direction. Accurate auditory stimulation and synchronization with the visual stream would provide an increment in the immersion sensation in these VR environments, which consequently could benefit the interaction between user and virtual audiovisual environment [89], [90], [91]. Nevertheless, and if this is true, what are the reasons for the historical predominance of visual IVEs? Why did auditory and, more importantly, audiovisual IVEs have, up to some point, been put on hold?

One of the main reasons that could help explain the predominance of investment (both financial and scientific) in visual IVEs is that there was an earlier understanding of some of the visual depth perception mechanisms. The systematic study and modeling of visual depth can be traced to the works of Filippo Brunelleschi in the Renaissance period [46] and to the artistic and scientific boom that followed the discovery of the perspective rules. The physical principles of these cues are well known and some are easily modeled with today's knowledge and technology (exceptions are, for instance, the convergence and accommodation depth cues, which are dependent on the user's eye behavior and therefore harder to control and simulate). On the other side, the study of auditory depth perception is relatively recent and, consequently, there is less compounded knowledge on the perceptual mechanisms that underlie the perception of auditory depth.

It is clear that humans are quite better at estimating sound location in a closed environment when compared with performance in open-field, and this fact led to the common idea that reflected sound should play a major role in sound localization [92]. This is indeed the case – one of the most powerful depth cues in auditory perception is the energy ratio of the direct and reflected sound [93] (see **Figure 29**).

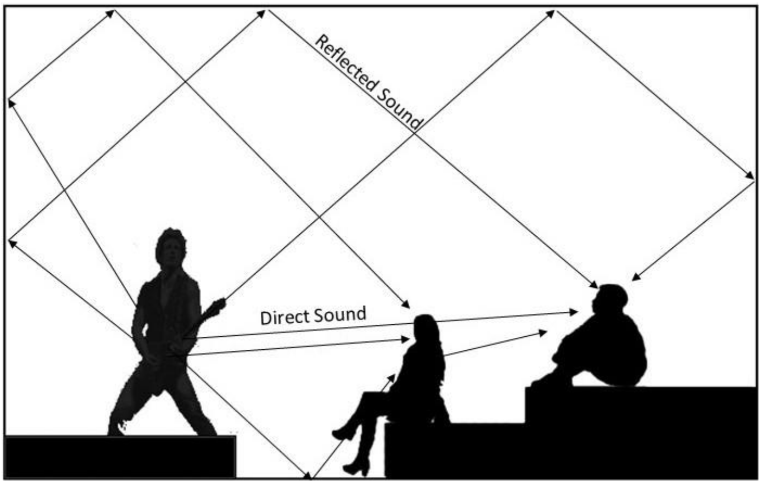


Figure 29. A depiction of how reflected sound works as a depth cue. A natural sound is emitted in an omni-directional way. The direct sound is the auditory signal that arrives at the listener without being reflect (and partly absorbed) by any physical structures around the emission site (walls, floor, ceiling), thus being the first and the more intense signal to arrive. It is the temporal and intensity differences between the direct and the reflected sounds that allow us to infer the location of the sound source.

In open-field, sound reflects only once (in the ground) and, as we can see by the work of Bronkhorst and Houtgast, we need around three to nine reflections orders (i.e. number of times that a sound signal is reflected on the room walls before its energy decrement render it inaudible) to accurately perceive sound distance [93]. In addition, the most important distance cue in an open-field scenario – loudness – is frequently erroneously perceived as the intensity level of the sound itself, which can cause misjudgments of stimulation distance. Other known important cues that allow us to locate sound in space are: sound pressure level (SPL) decrement with distance – which together with sound reflections are the most important information on **distance localization** –; and inter-aural time and level differences (ITD and ILD, respectively) normally referred to as *head related impulse response* (HRIR) or, more commonly, by its Fourier transform, the *head related transfer function* (HRTF) – which conveys information for **directional localization** (further description in **Section 3.4)**. The modeling of these sound features in order to accurately create virtual *spatialized* sound is not trivial, but several methods have appeared based on the well-known principles of ray tracing, bean tracing, boundary and finite element methods, and digital waveguide meshes. The most successful approach has been the Image Source Method (ISM) [94]. The popularity of this method is attested by its ever-increasing use in room acoustic modeling, in psychoacoustic investigation, in sound field analysis and synthesis, in sound rendering and auralization in virtual auditory systems, and in the design of acoustic spaces [95]. According to Lehmann and

Johansson, the prominence of the IMS technique can be attributed to a number of important benefits, when compared with other approaches [95], namely:

1) the simplicity of its algorithmic implementation;
2) an high degree of flexibility, with many simulation parameters (such as room dimensions, acoustic absorption coefficients, source and listener positions, and reverberation time and orders);
3) its ability to generate realistic room impulse responses (RIR) that are very similar to those obtained from real-room measurements;
4) the ability to investigate the effects of reverberation in isolation, separately from other sources of disturbance such as additive noise;
5) the guarantee to find all valid specular reflections in a given environment, which is critical when modeling the early part of a RIR where reflections (or the lack of them) can have a large effect on the perceived acoustic characteristics.

The software of sound rendering for use in VR environments is usually termed as an "auralization" software, which in the room-acoustic community means: "*the process of rendering audible, by physical or mathematical modelling, the sound field of a source in a space*." (p. 1) [96]. Auralization software, making use of the ISM in order to model the RIR of a sound source at a given position in relation to a defined listener, often uses a geometrical method for simulating the room reverberation. It simulates a sound impulse reflection by mimicking the reflected sound impulse with an external to the room sound source that gives the accurate time-delay and intensity decrement due to wall absorption (see **Figure 30**).

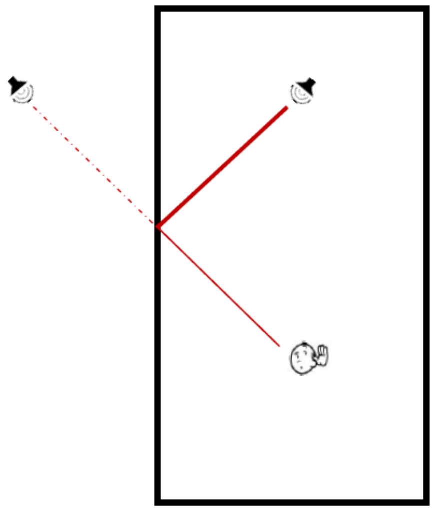


Figure 30. Graphical depiction of the ISM for sound reflection simulation. The room simulation in the ISM is just an indication on where the sound impulse should be mirrored and what intensity its reflection should have (taking into account the absorption rate of the wall). In fact, what the simulation does is to create innumerous sound sources around the room appointment and emitted the correct sound impulse considering what should have been the time and intensity absorption of the real room reflected sound.

Despite its success, the ISM has a huge drawback: a great computational cost it often required to render sound with several orders of reverberation[23]. The number of image sources necessary for the computation of a RIR by means of the ISM is proportional to the cube of the RIR length, and grows exponentially with the reflection order [95] (see **Figure 31**). These computational requirements make real-time applications of the ISM difficult to achieve in a satisfactory way.

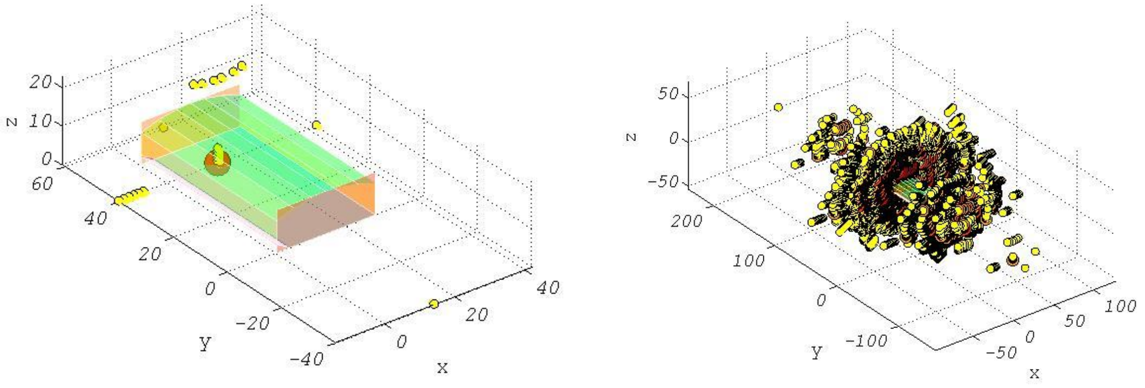


Figure 31.Graphical depiction of the external sound sources generated to simulate the wall reflected sound. In the left figure we have the sound sources generated in a one order reflection simulation, in the right we have the representation for a simulation with four orders of reflection. Images from a simulation performed with the auralization custom software currently in use at the Center for Computer Graphics, Guimarães.

This drawback is at the origin of one of the main problems with real-time auralization systems – the existence of end-to-end latency and the consequent lack of temporal and spatial accuracy of the virtual sound source. Thus, understanding both the user's temporal and spatial accuracy is paramount in order to suggest guidelines that help to mitigate the perceptual effects of this drawback.

In **Chapter 2,** we presented the theoretical background that justifies the use of PCMs as efficient user-research tools during the development of new UIs. We have made the case for empirical approaches that submit technology and technology users to controlled set-ups where valuable performance data from the latter can be used to improve the design of the former. We also made the case that this iterative logic (described in more detail in **Figure 4**), is quite suitable for the study and development of new IVEs and that is even more clearly understood when we analyze the limitations that current IVEs still present.

[23] Generating an IR of a sound source at 30m from the listener in a room with 24m width x 50m large x 11m high, took about 0.8 seconds using a one order reflection simulation, and about 3 hours using a 4 orders of reflection simulation – in a personal computer using a Intel® Core™ 2 Duo Processor P8600 (2,40GHz) with 8,00GB RAM and a operational system processing at 64 bits.

In the next chapter, we will apply some principles and methodologies described in this theoretical background and demonstrate five instances where we applied an empirical approach to study human perception and performance in IVEs. We will either contrast our findings with existing PCMs or we will propose new PCMs capable of explaining the perceptual or performance phenomenon under study. At the end of each experiment, a set of design guidelines are laid out, based on the results of each experiment.

3. EVALUATION OF HUMAN PERCEPTION AND PERFORMANCE AND DEVELOPMENT OF PREDICTIVE COGNITIVE MODELS

"Reality, however utopian, is something from which people feel the need of taking pretty frequent holidays"
Aldous Huxley, *Brave New World*

"The problem with the human observer is that he or she is human."
S. S. Stevens

In this chapter, we will present four different instances where controlled assessments of IVE's users where conducted with the common goal of assessing human perception and performance in IVEs and obtain sufficient knowledge to derive new PCMs and design guidelines that can be used to increase immersion or to test the fidelity of IVEs. These experiments vary regarding modality, stimuli complexity, and participant's task. Nevertheless, all of them where designed in order to function as the *Assess Users Experience of Interaction* tool of the framework presented in **Section 1.3** (**Figure 4**), and functioning as a vehicle of data that can be used to derive PCMs specially tailored to IVEs development.

In this perspective, all these experiments make use of empirical HCI techniques applied to the evaluation of user interface design (following the approach described in **Section 2.5**). Although there have been considerable technological advancements in the development of IVEs, there are still a lack of controlled methodologies to assess user experience and interaction and, consequently, a lack of PCMs dedicated to explain human perception and performance in IVEs. Thus, in this chapter we will be presenting highly-controlled environments and empirical methodologies design to measure perception and performance in projection-based IVEs and auralization systems, namely on:

- *Perception of audiovisual synchrony* – To understand how the *temporal* relation between auditory and visual stimuli (that might be affected by end-to-end delays) disturbs our perception of audiovisual synchrony.

- *Unity Assumption of audiovisual stimuli* – To understand how the *spatial* relation between auditory and visual stimuli (that might be affected by end-to-end delays) disturbs our perception of co-location of an audiovisual stimulus.

- *Visual distance perception* – To understand how to accurately measure visual distance perception in large projection screens and how to use this as a performance measure that can informs us about the IVE degree of fidelity;

- *Virtual sound location* – To understand how auditory location might be affected by the use of specific sound output equipment.

All the experiments described in this chapter were approved by the Direction Board of the Doctoral Program in Informatics, Department of Informatics, University of Minho. All participants gave their written informed consent. The experiments were conducted in accordance with the principles stated in the 1964 *Declaration of Helsinki*[24]*.*

All these experiment's implementations were made available (at https://github.com/Carlos-CCG/Thesis2019) and were developed in Open-source software, thus consisting an ideal tool in order to foster the development of PCMs in the IVEs research and development community.

3.1. Audiovisual Perception of Synchrony

3.1.1. On the Role of Temporal Coincidence

As we pointed out in **Section 2.1**, stimulating more than one human sense (also referred to as multimodality) is one of the essential characteristics of highly immersive virtual environments. Nevertheless, careless implementation of different modality actuation channels, without much consideration of how humans perceive multimodal scenes or without high control of the temporal relationship between the different output channels (e.g. image projection and sound output), can in fact have the opposite effect.

To the more general phenomenon of perceiving bimodal stimulus as such, we call it *Unity Assumption* (UA) (mostly studied with visuo-tactile an audio-visual stimulation) [97]–[99]. The UA theory states that a stimulus is perceived as a multimodal audiovisual stimulus, for instance, "... if the two sensory modalities are providing information about a sensory situation that the observer has strong reasons to believe (not necessarily consciously) signifies a single (unitary) distal object or event." [100]. Furthermore, the level of certainty with which we perceive an object or event as multimodal appears to

[24] World Medical Association Declaration of Helsinki – Ethical Principles For Medical Research Involving Human Subjects. Can be downloaded here: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

be positively correlated with the number of *amodal* stimulus properties shared by the two sensory modalities. Amodal properties are physical properties that might be shared by different perceptual modalities, such as: *spatial location*, *temporal pattern*, *size*, *shape*, *orientation*, *intensity*, *motion*, *texture* [100]. In particular, audio-visual stimuli appear to be best perceived as a unitary audiovisual stimulus when both modalities present [99]:

- Temporal coincidence;
- Spatial coincidence;
- Motion vector coincidence;
- Causal determination;

In IVEs, these first two amodal properties – temporal coincidence and spatial coincidence – are of special relevance, particularly when we are dealing with systems with measurable end-to-end delays and targeted at applications requiring precise temporal and spatial fidelity. Thus, in this section we will explore the role of temporal coincidence, while in **Section 3.2** we will explore the role of spatial coincidence.

Humans are especially sensitive to audio-visual temporal relations, and thus in order to accurately implement an audiovisual IVE we must play particular attention to the aspect of audiovisual synchrony. Guidelines on how to measure, control, and set-up the appropriate temporal relation between auditory and visual stimuli of an IVE are dependent on knowledge about the fundamentals of human audiovisual synchrony perception – a phenomenon that is far from being clear and well understood [44].

Humans face intricate problems to perceive synchrony in audiovisual real-world events because of the relative timing of visual and auditory inputs. These problems are related to both physical and neural differences underlying sound and light propagation and human processing of these stimuli. When a natural audiovisual event occurs, the visual and auditory signals are synchronic at the origin, since they are caused by the same physical source and thus emitted at the same time. However, there are significant differences in the propagation time for light and sound, as sound takes about 3.4ms to travel 1 meter[25] and light travels approximately 299 792 meters in 1ms. Nevertheless, in our daily life and within a certain distance range, audiovisual stimuli are still perceived as synchronic at the source. Given the abovementioned sound delay (in relation to light) of about 3ms per meter traveled from the perceived

[25] The standard value for the speed of sound (343 meters per second) is calculated as its value at 20°C and traveling in the elastic medium of air. Nevertheless, this value is dependent from both the temperature and the type of elastic medium through which the sound is been propagated.

object, it is not obvious how certain audiovisual stimuli are perceived by the user as an audiovisual unitary phenomenon.

These constrains make the problem of audiovisual synchronization especially relevant for the development of audiovisual IVEs, mainly because common range distances from stimuli sources to observers are often large enough to create a considerable gap between the arrival times of visual and auditory signals in the real-world. Thus, when developing an audiovisual IVE, one should know if accurately modeling of this variable audio-image delay is relevant or not. Although physical realism would advise us to consider the slower propagation time of auditory stimuli, in terms of computational cost it would be unnecessary to do so if users are not sensitive to these delays. Moreover, implementing an audio delay based on the simulated distance of the audiovisual stimuli requires not only adaptation of the auditory and visual scene in the VR platform, but also complex measurements of each system (hardware) output channels. In order to know the audio-visual temporal relation of the output channels, we need to measure it externally, resorting to light and acoustic sensors, since the end-to-end delay from the video and the audio channels vary as a function of the graphic board, the sound card, the tracking system, or even the defined projection frame-rate [32].

Moreover, differences in propagation time are not the only aspect to consider in audiovisual synchronization. Humans also present differences at the level of transduction times of the human perceptual channels [101] [102], only this time with sound being transduced faster (~ 1 ms; see [103]) than light (~50 ms). However, as these neural temporal differences are constant, observers become adapted to this difference due to a long history of exposure to such "veridical" neural lags [104], due to a phenomenon called *perceptual temporal recalibration* [105]. Nevertheless, it remains to be explained how can we perceive audiovisual events as unitary when the audio and the visual streams will physically arrive to the observer at variable asynchronies as a function of the distance of presentation.

Several studies have shown that audiovisual integration does not require temporal alignment between the visual and the auditory stimuli. We still perceive as synchronic visual and auditory stimuli that are not received or emitted at the same time (e.g. [106]–[112]). Notwithstanding, temporally mismatched stimuli can only be perceived as synchronic when keeping the onset difference between sound and image within certain temporal limits, termed *Window of Temporal Integration* (WTI) [44] [113]. In multisensory perception, this phenomenon can be defined as the range of temporal differences on the onset of two or more stimuli of different modalities where these are still best perceived as a unitary multisensory stimulus. As Vroomen and Keetels [44] pointed out, the main reason why signals from different sensory modalities are perceived as being synchronic despite these differences, is that the brain judges as synchronic two stimulation streams that arrive within a certain amount of temporal disparity. Thus, another important piece of information for IVEs developers would be to know what are the limits of audiovisual WTI and how those relate with the distance of the audiovisual scene. By knowing this, they should aim at keeping systems end-to-end delays of each modality inside the values tolerated by the WTI.

Research on this phenomenon has provided us with surprising findings. A large number of studies on audiovisual temporal alignment have found that we perceive stimuli from different modalities as being in maximal synchrony if the visual stimulus arrives at the observer shortly before the auditory stimulus (e. g., [106], [109], [110], [114]). This finding has been termed the *vision-first bias* [113] [44]. In a work that boosted the scientific discussion on the vision-first bias, [110] used a *temporal-order judgment* (TOJ) psychophysical task (see **Figure 32**) to assess the perceived temporal relation between the emission of a sound (a burst of white noise) and a brief light flash. The flashes of light were displayed from LEDs located at distances of 1, 5, 10, 20, 30, 40 and 50 m from the participant. The sound was always transmitted by headphones but was compared with the visual stimuli at different distances. In their paper published in *Nature*, Sugita and Suzuki reported that the *Stimulus Onset Asynchrony* (SOA) that provides the best perception of synchrony is always a positive one (i.e., sound lagging image) and most importantly, when the distance of visual stimuli increases, larger sound lags are observed at the *Point of Subjective Simultaneity* (PSS) (i.e., the temporal relation between sound and image which elicits the higher percentage of "synchronous" answers, see **Figure 32**). Their results are roughly consistent with the velocity of sound, at least up to 20 m of visual stimulus distance, and can be quite well predicted by a linear model based on this physical rule. Thus, and according to [110], it seems that the brain considers sound propagation velocity when judging synchrony, relying on distance information to compensate for the natural differences on the stimuli's propagation velocity. Other studies have also pointed to a perceptual mechanism of compensation for differences in propagation velocity (e.g., [106], [109], [115]), further suggesting that we resynchronize the signals of an audiovisual event by shifting our PSS in the direction of the expected audio lag.

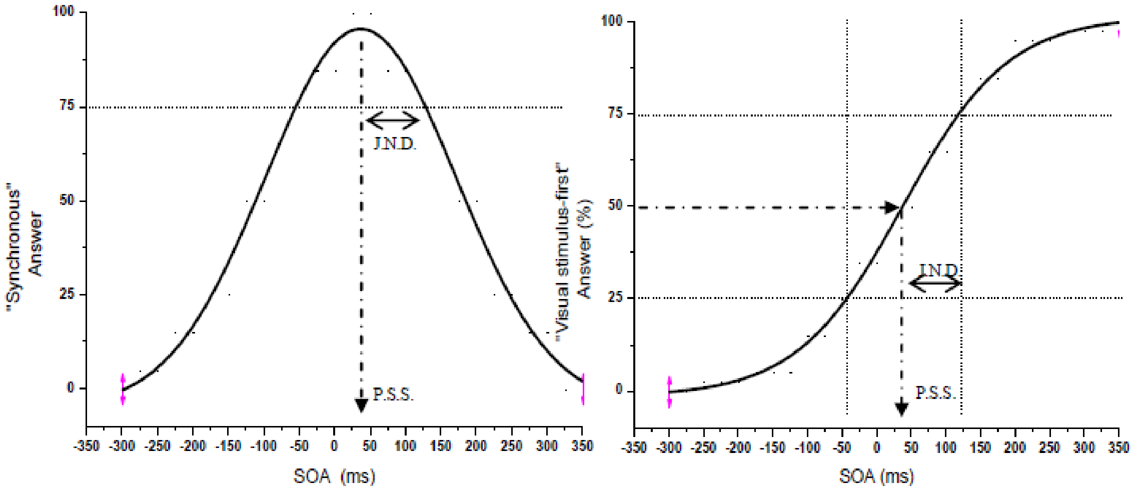


Figure 32. Graphically representation of the PSS and the WTI (corresponding to the Just Noticeable Difference – JND) on a Simultaneity Judgement (SJ) Task – where the question posed to the participant is of the type: "is the presented audiovisual stimulus synchronous regarding visual and auditory streams?" – and on a Temporal Order Judgment (TOJ) task – where the question posed to the participant is of the type: "what was the stimulus presented first, the visual or the auditory one?"

Nonetheless, other researchers failed to observe such compensatory effect [116], [117]. Lewald and Guski [116] tried to replicate the findings of Sugita et. al [110] in a less artificial setting. Using the same kind of stimuli (sound bursts and LED flashes) but co-located, they found no distance compensation. In fact, the PSS shifted in the opposite direction. In this experiment, participants had the best perception of synchrony when auditory and visual signals were synchronic in their arrival at the observer's sensorial receptors (i.e. sound leading image on onset). As Lewald and Guski pointed out, "this conclusion is in diametral opposition to the study of Sugita and Suzuki" (p. 121). According to these authors, there are some limitations in Sugita and Suzuki's study that might have been affecting the ecology of the stimuli (i.e., ecology in this sense refers to the capacity for the stimuli to simulate the real world). Firstly, the sound stimuli were not co-located with the visual stimuli and consequently there was no auditory distance information. Secondly, the luminance of the visual stimuli was increased to compensate for the light intensity attenuation with distance and, by doing this, they kept the perceived stimuli's luminance constant, thus providing a cue that is incongruent with the expected information on distance increment.

Nevertheless, both the studies of [98] and [99] also present potential problems in the simulation of distance, namely:

1. By conducting the experiment in open-field, Lewald and Guski failed to provide the optimal conditions for auditory distance information. One of the most powerful auditory depth cues is

the ratio of energies of direct and reflected sounds [93], which is almost absent in open-field stimulation. Also, the most important depth cue in a situation of open-field – loudness – is frequently and erroneously perceived as the level of the sound itself in the absence of relative loudness cues, which can cause misjudgments of stimuli distance;

2. The use of artificial stimuli, such as flashes and beeps, eliminated two other relevant depth cues: familiar size of the visual stimuli and familiar loudness of the auditory stimuli.

3. Arnold and colleagues manipulated the angular size and velocity (i.e. retinal size and velocity) of their visual stimuli to ensure that the size and velocity of the stimuli appeared constant while distance increased. This poses the same problem of incongruent depth cues pointed out by Lewald [116] to the work of Sugita [110].

Additionally, there are limitations that are common to both groups of studies. Recent findings show that in the study of audiovisual synchrony, biological motion stimuli are preferable over rigid motion or stationary stimuli [107], [112]. For instance, the use of point-light displays with biological motion has been previously pointed out as an important factor in simultaneity judgment tasks. In a study where participants had established a baseline PSS to an audiovisual stimulus consisting of footage of a professional drummer playing a conga drum, Arrighi and collaborators found that the PSS was not different when the visual stimulus was a computerized abstraction of the drummer with the same movement (biological) presented in the footage. However, when the same artificial visual stimulus had an artificial motion pattern (constant velocity) the PSS was significantly different, even when the frequency was the same in both conditions [107]. Petrini, Holt, and Polick have found that a non-natural orientation of a point-light drummer can affect the simultaneity judgment of non-musical expert participants, thus showing that naturalistic representations are preferred in this kind of tasks [118].

3.1.2. Assessing Audiovisual Synchrony in an IVE

Considering the above controversy and taking into account the main critiques regarding previous studies on compensatory mechanisms, in this experiment we report findings that offer a clearer answer to the following questions:

1 How can we perceive an audiovisual event as such, if the audio and the visual signals physically arrive at different times to the observer?

2 Does distance play a role in a possible mechanism of perceptual compensation?

3 What are the levels of audiovisual asynchrony that users of IVEs can cope with without clearly perceiving an audiovisual mismatch?

In order to answer these questions, we developed an audiovisual synchrony judgment task, capable of being performed in an IVE, and resorting to audiovisual biological motion stimuli. Biological stimuli such as PLW also allow for high levels of control and manipulation of critical parameters for visual and sound depth perception. Namely, angular and familiar size, angular velocity, elevation, intensity, contrast, and perspective for visual distance judgments; and sound pressure level, high frequency attenuation, and reflected sound for auditory distance judgments. Thus, in order to assess the existence of a mechanism that compensates for the differences in propagation velocity we presented the audiovisual biological stimulus at several distances in a controlled environment simulating the depth conditions of the real physical world. Additionally, by manipulating the number of depth cues presented in the stimuli we were able to assess if a perceptual mechanism that compensates for audiovisual asynchronies does exist and if it has a relation with distance of presentation of the audiovisual stimulus.

In short, we expect to support the argument of perceptual compensation if we find a positive relation, close to the rule of physical propagation of sound, between the shift of the PSS (towards an audio lag) and the distance of the stimuli. We also expect this relation to be dependent on the number and quality of the depth cues. The results from this experiment will be discussed in light of their implications for the conceptualization of the mechanisms of simultaneity constancy and development of audiovisual IVEs.

Method

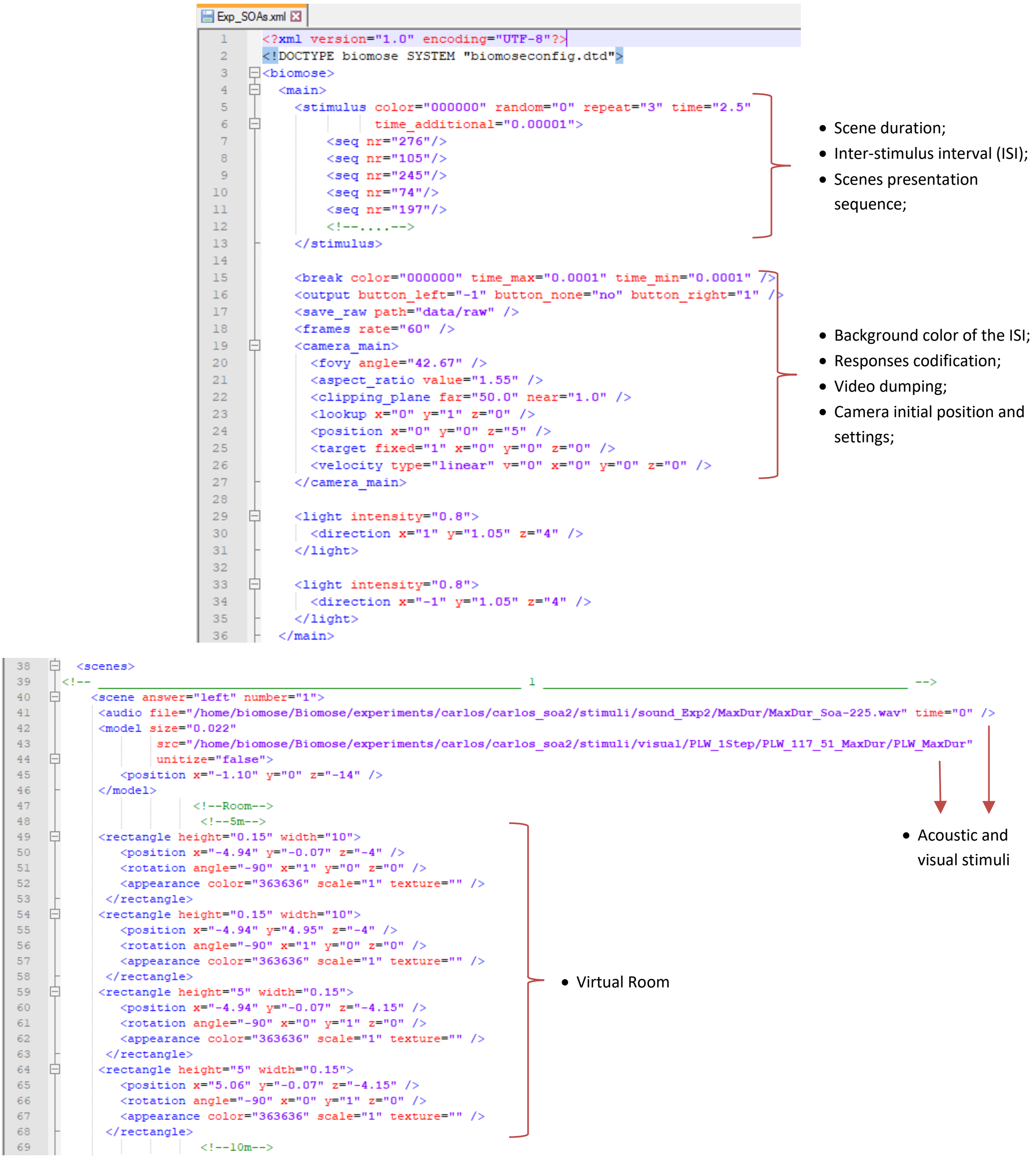
Participants: Four participants, aged 21-33 years old, underwent visual and auditory standard screening tests and had normal hearing and normal, or corrected to normal, vision. All were voluntary university students or researchers, and all gave written informed consent to participate in the study. Two had some background knowledge about the thematic of the study and the remaining two were naïve as to the purpose of the experiment.

Stimuli and Materials: The experimental tasks were performed in a darkened room at the Center for Computer Graphics in the University of Minho. For the stimuli presentation, we used a cluster of PCs with NVDIA® Quadro FX 4500. A 3chip DLP projector Christie Mirage S+4K with a resolution of 1400x1050 pixels and a frame rate of 60Hz was used for the projection of the visual stimuli. The area of projection was 2.80 m high and 2.10 m wide. The sound was presented using a computer with a Realtek Intel 8280 IBA sound card through a set of Etymotics ER-4B in-ear phones. Latencies between visual and auditory channels were measured and adjusted using a custom-built latency analyzer consisting of an Arm7 microprocessor coupled with light and sound sensors.

We programed this experiment using the XML-based BioMoSe[26] language (see **Example 2**). BioMoSe (*Biological Motion Segmentation*) is a custom-made software, developed at the Laboratory for Visualization and Perception of the University of Minho, which uses the *Open Graphics Library* (OpenGL) to control and parametrize the displaying of the visual and auditory stimuli. As a virtualization platform, we used *VR Juggler* [119], an Open Source platform for virtual reality applications developed by the Cruz-Neira research group at Iowa University and thus specially tailored for controlled displaying of audiovisual stimuli in CAVE-like environments. VR Juggler is particularly useful for multiscreen applications, since it allows us to configure the correct partitioning between projection nodes and accurate edge blending in order to avoid bad continuity projection problems.

Additional materials used in the implementation of this experiment can be found at:
https://github.com/Carlos-CCG/Thesis2019/tree/master/ExpSynchrony

[26] *Extensible Markup Language*



```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE biomose SYSTEM "biomoseconfig.dtd">
<biomose>
  <main>
    <stimulus color="000000" random="0" repeat="3" time="2.5"
              time_additional="0.00001">
        <seq nr="276"/>
        <seq nr="105"/>
        <seq nr="245"/>
        <seq nr="74"/>
        <seq nr="197"/>
        <!--....-->
    </stimulus>

    <break color="000000" time_max="0.0001" time_min="0.0001" />
    <output button_left="-1" button_none="no" button_right="1" />
    <save_raw path="data/raw" />
    <frames rate="60" />
    <camera_main>
      <fovy angle="42.67" />
      <aspect_ratio value="1.55" />
      <clipping_plane far="50.0" near="1.0" />
      <lookup x="0" y="1" z="0" />
      <position x="0" y="0" z="5" />
      <target fixed="1" x="0" y="0" z="0" />
      <velocity type="linear" v="0" x="0" y="0" z="0" />
    </camera_main>

    <light intensity="0.8">
      <direction x="1" y="1.05" z="4" />
    </light>

    <light intensity="0.8">
      <direction x="-1" y="1.05" z="4" />
    </light>
  </main>
```

- Scene duration;
- Inter-stimulus interval (ISI);
- Scenes presentation sequence;

- Background color of the ISI;
- Responses codification;
- Video dumping;
- Camera initial position and settings;

```xml
  <scenes>
<!-- ____________________________________ 1 ____________________________________ -->
    <scene answer="left" number="1">
     <audio file="/home/biomose/Biomose/experiments/carlos/carlos_soa2/stimuli/sound_Exp2/MaxDur/MaxDur_Soa-225.wav" time="0" />
     <model size="0.022"
             src="/home/biomose/Biomose/experiments/carlos/carlos_soa2/stimuli/visual/PLW_1Step/PLW_117_51_MaxDur/PLW_MaxDur"
             unitize="false">
        <position x="-1.10" y="0" z="-14" />
     </model>
                 <!--Room-->
                  <!--5m-->
     <rectangle height="0.15" width="10">
        <position x="-4.94" y="-0.07" z="-4" />
        <rotation angle="-90" x="1" y="0" z="0" />
        <appearance color="363636" scale="1" texture="" />
      </rectangle>
     <rectangle height="0.15" width="10">
        <position x="-4.94" y="4.95" z="-4" />
        <rotation angle="-90" x="1" y="0" z="0" />
        <appearance color="363636" scale="1" texture="" />
      </rectangle>
     <rectangle height="5" width="0.15">
        <position x="-4.94" y="-0.07" z="-4.15" />
        <rotation angle="-90" x="0" y="1" z="0" />
        <appearance color="363636" scale="1" texture="" />
      </rectangle>
     <rectangle height="5" width="0.15">
        <position x="5.06" y="-0.07" z="-4.15" />
        <rotation angle="-90" x="0" y="1" z="0" />
        <appearance color="363636" scale="1" texture="" />
      </rectangle>
                  <!--10m-->
```

- Acoustic and visual stimuli

- Virtual Room

Example 2. Biomose declaration of the main settings of the experimental scenario (Top panel); XML declaration of the scene visual and auditory stimuli (Bottom panel).

Three experimental conditions differing from each other in the number of depth cues were presented. The "audiovisual depth cues" condition presented visual and auditory depth cues coherent with the simulated presentation distance. In the "visual depth cues" condition only the visual depth cues varied coherently with the simulated presentation distance, but sound had no distance cues (anechoic sound without room-acoustic information). The “reduced depth cues” condition presented both visual and auditory depth cues impoverished.

A) Auditory Stimuli.

In the “audiovisual depth cues” condition the auditory stimulus consisted in a binaural sound recorded using a Brüel&Kjaer® head and torso simulator Type 412SC. An anechoic step sound was emitted trough a Brüel&Kjaer® omnidirectional Loudspeaker Type 4295 inside a sports pavilion 24 m wide, 50 m long and 11 m high. The step sound was recorded at distances of 10, 15, 20, 25, 30, and 35 m from the head and torso simulator, located 6 m from one end of the sports pavilion (see **Figure 33**). The anechoic step sound came from the database of controlled recordings from the College of Charlston [120] and corresponds to the sound of a male walking over a wooden floor and taking one step. In the two remaining conditions, the auditory stimulus was an auralized sound of the above-referred anechoic recording, with directional cues matching the visual stimulus, but in free field (thus without any room-acoustic information) and without air attenuation with distance.

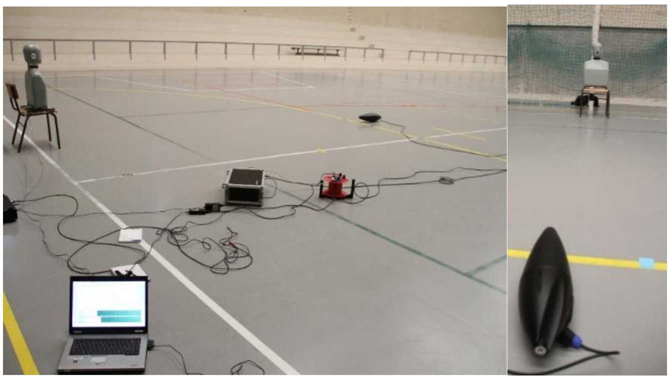


Figure 33. Set-up for acquisition of the step sounds binaural recordings. We used an omnidirectional speaker (Bruel&Kjaer Type 4295) to emit the anechoic step-sounds and an Head & Torso simulator (Bruel&Kjaer) to record the binaural stimuli.

B) Visual Stimuli.

Different visual stimuli according to the experimental condition were presented to participants. In both the “audiovisual depth cues” and the “visual depth cues” conditions the visual stimuli were Point-Light Walkers (PLW) moving in a front-parallel plane to the observer and taking one step aligned with the center of projection. The PLW was walking inside a simulated room with the same dimensions of the sports pavilion referred before. The PLW was composed of 13 white dots (54 cd/m2) that moved against a black background (0.4 cd/m2) and was generated in the Laboratory of Visualization and Perception (University of Minho) from motion data captured using a Vicon® system with 6 MX F20 cameras and a set of custom LabVIEW routines. All stimuli corresponded to the correct motion coordinates of a 1.78 m high, 17 year-old male, walking at a velocity of 1.1 m/s. The duration of the visual stimulus was variable from scene to scene in order to avoid the use of a fixed time between the beginning of the scene's projection and the occurrence of the step. Thus, there were three different stimulus durations: 1.08 ms, 1.12 ms and 1.17 ms; and the PLW step occurred at the 527th ms in the minimum duration; at the 547th ms in the medium duration; and at the 569th ms in the maximum duration (**Figure 34**).

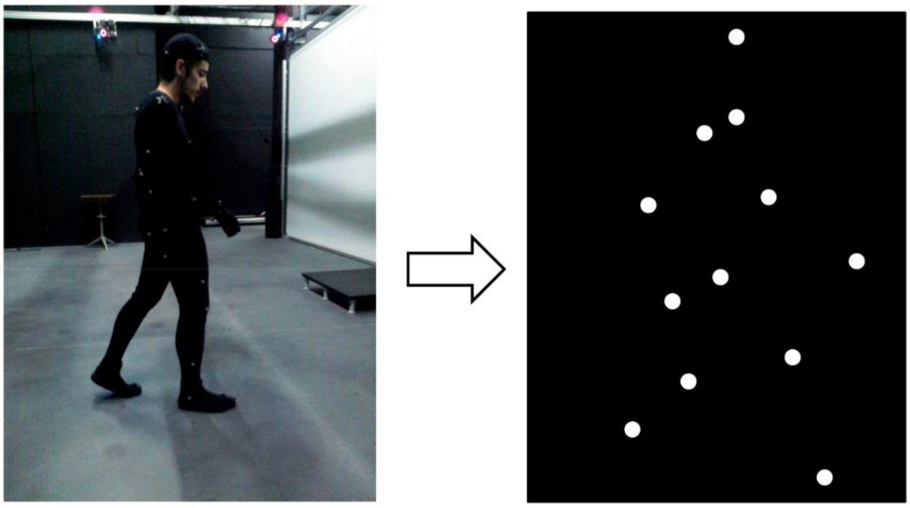


Figure 34. Point Light Walkers (PLWs) Capture. On the left panel a frame of a participant in a session of biological motion capture is presented. The markers placed on the suit in specific locations, are designed to reflect near infrared light transmitted by the 6 cameras. This way we could record in real time the position (along 3 axes) of each marker in order to design an animated representation of the human body movement (image on the right panel).

To simulate the six stimuli distances (10, 15, 20, 25, 30 and 35 m), the visual angular size and angular velocity of the stimuli was changed according to the expected changes in the physical world.

Using PLWs, two important pictorial depth cues were made available: familiar size and elevation. Furthermore, these visual stimuli also presented two dynamic depth cues: the amplitude of the step (wider steps represent a closer presentation) and the angular velocity (a smaller angular velocity translates into a farther distance of presentation). The PLW allowed the coordination of these depth cues by decreasing the angular size and velocity, as well as by decreasing the angular size of the dots composing it, and by gradually increasing its elevation according to the stimuli distance.

The room simulation was developed using perspective depth frames. The floor and wall lines were virtually located at 10, 15, 20, 25, 30, and 35 m from the observer. Thus, the PLW corresponding to each one of these distances was presented as walking on top of the correspondent ground line (**Figure 35**).

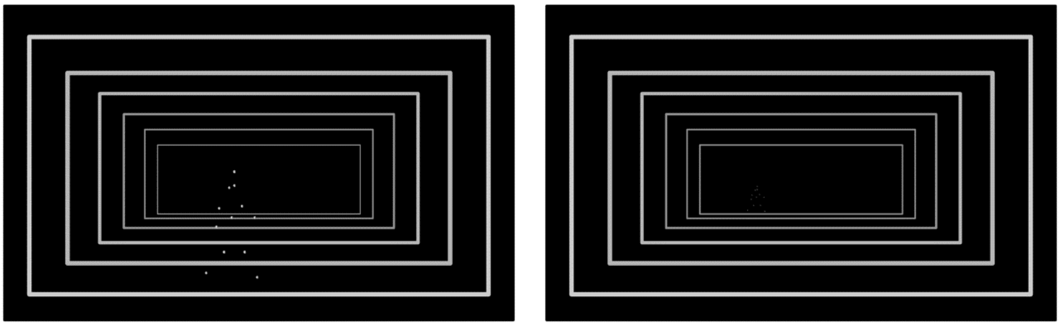


Figure 35. Perspective Depth Cue. A video of the stimuli used can be visualized here: https://figshare.com/articles/_Depth_Cues_and_Perceived_Audiovisual_Synchrony_of_Biological_Motion_/851388

In the "reduced depth cues" condition, most depth cues were eliminated from the visual stimuli. These stimuli presented no room perspective cues since only the feet were presented, with a random size of dots and at a constant elevation. This way we also eliminated the bodily cues of familiar size. Thus, the single depth cue available was the amplitude of the step, with wider steps meaning a closer presentation.

C) Visual and Auditory Stimuli Relation.

In order to present several audiovisual events, the visual and the auditory stimuli were combined in 19 different stimuli onset asynchronies (SOAs) for several distances (**Figure 36**). The SOAs took the values displayed in **Table 5**:

Table 5. SOAs values for the difference presentation distances. Negative values indicate that sound step was presented before the visual feet touched the ground, and positive values indicate that sound was presented after the visual step.

PLW distance	SOAs
Even distances: 10, 20, 30 meters	-240 ms; -210 ms; -180 ms; -150 ms; -120 ms; -90 ms; -60 ms; -30 ms; 0 ms; 30 ms; 60 ms; 90 ms; 120 ms; 150 ms; 180 ms; 210 ms; 240 ms; 270 ms; 300 ms;
Odd distances: 15, 25, 35 meters	-225 ms; -105 ms; -165 ms; -135ms; -105 ms; -75 ms; -45 ms; -15 ms; 0ms; 15 ms; 45 ms; 75 ms; 105 ms; 135 ms; 165 ms; 195 ms; 225 ms; 255 ms; 285 ms.

The reason why we had a different spectrum of SOAs for even and odd distances was twofold: Firstly, we wanted to always present the theoretical values that, according to the compensation for sound velocity hypothesis, would provide the best sensation of audiovisual synchrony (e.g., 30ms for stimuli at 10m from the observer, 45ms at 15m, and so on). Secondly, we wanted to keep the experimental sessions within a reasonable duration to avoid fatigue effects. Therefore, we had two groups of 19 SOAs presented, each one at 3 distances comprising a total of 114 different audiovisual stimuli.

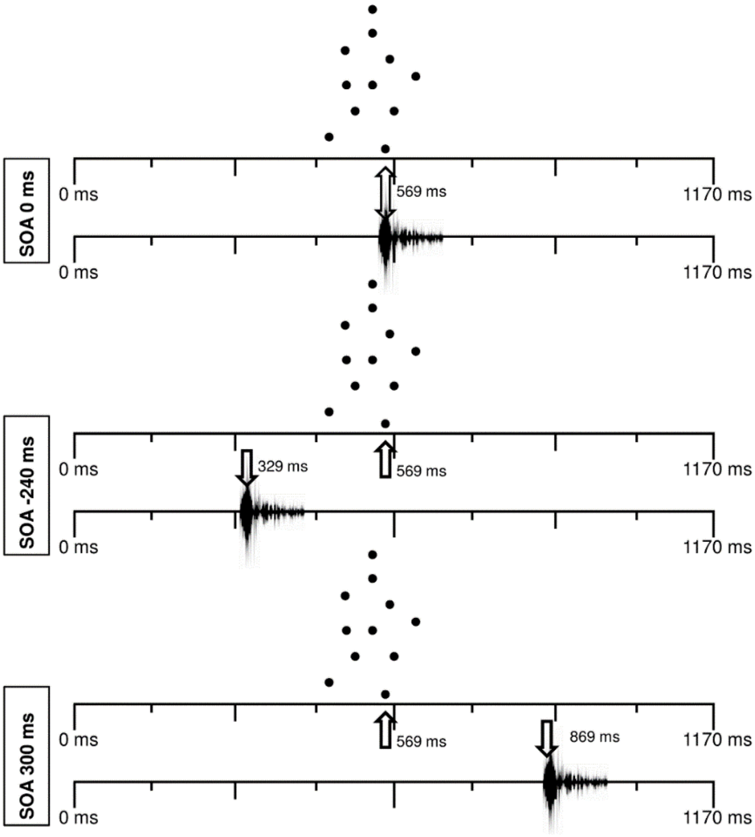


Figure 36. Examples of audiovisual stimuli where we can see when steps occur. We can see the temporal relations between the occurrence of the visual step and the occurrence of the auditory step in the SOAs 0ms, -300ms, and +300ms.

Procedure: In each experimental session, we presented the PLW at 3 different distances in a random order. All stimuli were randomly presented with 40 repetitions and with a given duration according to the condition: 5.5s in the "audiovisual depth condition" and 2.5s in the "visual depth condition" and in the "reduced depth condition". There was an inter-stimuli interval of 1.5s for all conditions. Before each experimental session, the participants were shown 10 repetitions of an audiovisual stimulus in which the sound appeared with a 300ms lead, and 10 other repetitions of an audiovisual stimulus in which the sound appeared with a 330ms lag. This preliminary session was taken in order to check if participants were able to perceive any kind of asynchrony. None of these SOAs used in this preliminary session were then used in the experimental session.

At the beginning of the experimental session the following instructions were given: "You are going to participate in a study in which you will be presented with several audiovisual scenes of a PLW walking at a certain distance. I want you to pay close attention to the audiovisual scene, because you will have to judge its audiovisual synchrony. In this scene you will see a walker taking one step and you will hear his step sound. The distance of presentation may vary between 10, 20, and 30m (or 15, 25, and 35m, in some trials). After each scene, if you think that the auditory and the visual streams were synchronized click the right button; otherwise, if you think that the auditory and the visual streams were not synchronized click the left button. Please give your answer only after the visual and auditory stimuli presentation". The participant was seated in a chair 4 m from the screen and in line with the center of the projection area. In each scene the participant was presented with a PLW walking from left to right and taking one step at a velocity of 1.1m/s, while listening trough in-ear phones to one step in a given temporal relation with the visual stimulus. After the presentation of each audiovisual stimulus and during the inter-stimulus interval, the participant had to answer in a two-key mouse according to the instructions. The experimental sessions were blocked by condition and the conditions' order was randomized between participants.

Results: The individual analyses of PSS and WTI for all participants are shown in **Table 6**. PSSs were obtained by adjusting a Gaussian function to the data, of the type:

Equation 9. $y = y_0 + A * e^{-0.5\left[\frac{x-\mu}{\sigma}\right]^2}$ *, where A is the curve amplitude (difference between the highest and the lowest value of the distribution in the y-axis), y0 is the lowest value of the distribution in the y axis, σ^2 is the variance and (µ, y0+A) is the distribution's peak coordinates.*

WTIs were calculated following the method presented in the review of Vroomen and Keetels [44].

Table 6. Individual values of the PSS and WTI. The values are presented in ms and for the several distances in each stimulus condition. In the last column the equations and the values of adjustment for each of the linear functions fitted to the individual data are presented. The asterisks signal the participants that had some background knowledge about the thematic of the study.

Part.	***Condition***	***PSS 10m (WTI)***	***PSS 15m (WTI)***	***PSS 20m (WTI)***	***PSS 25m (WTI)***	***PSS 30m (WTI)***	***PSS 35m (WTI)***	***Linear Fitting***	***Adjust. (R^2)***
1*	"Audiovisual Depth Cues"	16 (77)	2 (73)	18 (80)	26 (88)	60 (100)	57 (84)	y = 2.21x - 19.9	0.71
	"Visual Depth Cues"	54 (98.5)	69 (113)	69 (106)	80 (96.5)	113 (78)	95 (86.3)	y = 1.96x + 35.62	0.71
	"Reduced Depth Cues"	89 (108)	83 (117)	86 (98.5)	62 (88)	75 (86.5)	61 (79.5)	y = -1.07x + 100.3	0.60
2	"Audiovisual Depth Cues"	-10.5 (102)	-13 (89)	30 (104)	36 (138)	74 (100)	91 (106)	y = 4.42x - 64.9	0.93
	"Visual Depth Cues"	35 (86)	43 (70)	47 (77)	52 (69.5)	43 (87)	71 (60.5)	y = 1.04x + 24.88	0.54
	"Reduced Depth Cues"	46 (66)	31 (69)	40 (60.5)	27 (62)	10 (49.5)	18 (68)	y = -1.20x + 55.50	0.66
3*	"Audiovisual Depth Cues"	22 (126)	-3 (118)	28 (114)	20 (117)	75 (105)	84 (125)	y = 3.06x - 31	0.62
	"Visual Depth Cues"	41 (107)	40 (109)	45 (104)	47 (116.5)	59 (90.5)	63 (116.5)	y = 0.97x + 27.16	0.85
	"Reduced Depth Cues"	2 (116)	-11 (115)	-4 (113.5)	-13 (121)	-18 (109)	5 (117)	y = -0.07x - 4.88	0.42
4	"Audiovisual Depth Cues"	50 (253)	69 (232)	68 (161)	76 (213)	106(124)	126(160)	y = 2.86x + 18.23	0.88
	"Visual Depth Cues"	103 (132)	88 (106)	117(125)	108 (119)	96 (101)	100(100.5)	y = 102	0.71
	"Reduced Depth Cues"	-12 (91.5)	-13 (103)	-25 (91)	-30 (64.5)	-19 (101)	-13 (91)	y = -0.17x - 14.73	0.71

Two participants had some background knowledge about the thematic of the study and the remaining two were naïve to the purpose of the experiment. According to independent sample t-tests there was no significant difference between both groups with regard to the PSS values in the "reduced depth cues" condition (t (22) = 1.99, n.s.)[27], in the "visual depth cues" condition (t (22) = -0.98, n.s.) and in the "audiovisual depth cues" condition (t (22) = -1.68, n.s.). Regarding the WTI values, there were no significant differences among participants regarding their background knowledge about the study's thematic in the "visual depth cues" condition (t (22) = .962, n.s.). There were, however, significant differences regarding the WTI values in the "reduced depth cues" condition (t (22) = 4.4, p<.01) and in the "audiovisual depth cues" condition (t (22) = -2.78, p<.05). Participants with background knowledge had higher values of WTI in the "reduced depth cues" condition, but lower values of WTI in the "audiovisual depth cues" condition.

[27] The notation of an hypothesis test is presented here according to the following rules (for detailed information see [237]):

- The first letter indicates the type of statistical test – *t* for the Student's T-Test (independent or paired samples) and *F* for Analysis of Variance;
- The numbers in brackets indicate the degrees of freedom (d.f.) of the statistical test – in a Student's T-Test the d.f. is equal to *n-1* (n° of observations minus 1); in an Analysis of Variance, two d.f. values are reported, the d.f. from the Model (n° of groups minus one, also denoted as k-1) and degrees of freedom from unsystematic variation of the data (the total sample size minus the number of groups, also denoted as N-k);
- The value of the test, in the case of a T-Test is a mean and for an ANOVA is a variance
- The p-value, also called the significance value, is a probability value that tell us if our model fits the data well. If p-value is less than .05 (significant), we generally reject the null hypothesis. If p-value greater than .05 (or non-significant, represented with n.s.), then the effect is most likely to occur by chance.

Since significant differences in the PSS between both groups were not found, a One-way analysis of variance (ANOVA) was calculated for all participants' PSS values. This revealed differences between the participants' PSS values for the "reduced depth cues" condition (F (3,20) = 93.16, p<.01), the "visual depth cues" condition (F (3,20) = 20.43, p<.01) and the "audiovisual depth cues" condition (F (3,20) = 3.32, p<.05). The Scheffé Post-hoc tests show that these differences between participants are significant in all the comparisons of the "reduced depth cues" condition (with the exception being the comparison between participant 3 and participant 4) and in some comparisons of the "visual depth cues" condition, where participant 1 has a significantly higher PSS than participant 2 (p<.01), and participant 4 has a PSS significantly higher than participant 2 (p<.01) and participant 3 (p<.01). The individual analyses of PSS and WTI for all participants show similar response patterns across participants (i.e., the slope of the linear fitting to the PSS values increases when we go from conditions with less depth cues to conditions with more depth cues). Therefore, we chose to pool all the individual data for a more detailed analysis.

Figure 37, shows the fitting of a Gaussian function to the data pool for distances of 10, 20, 30 m (graph on the left) and 15, 25, 35 m (graph on the right), for each one of the conditions. All the data, grouped by distance, conformed well to the Gaussian fittings.

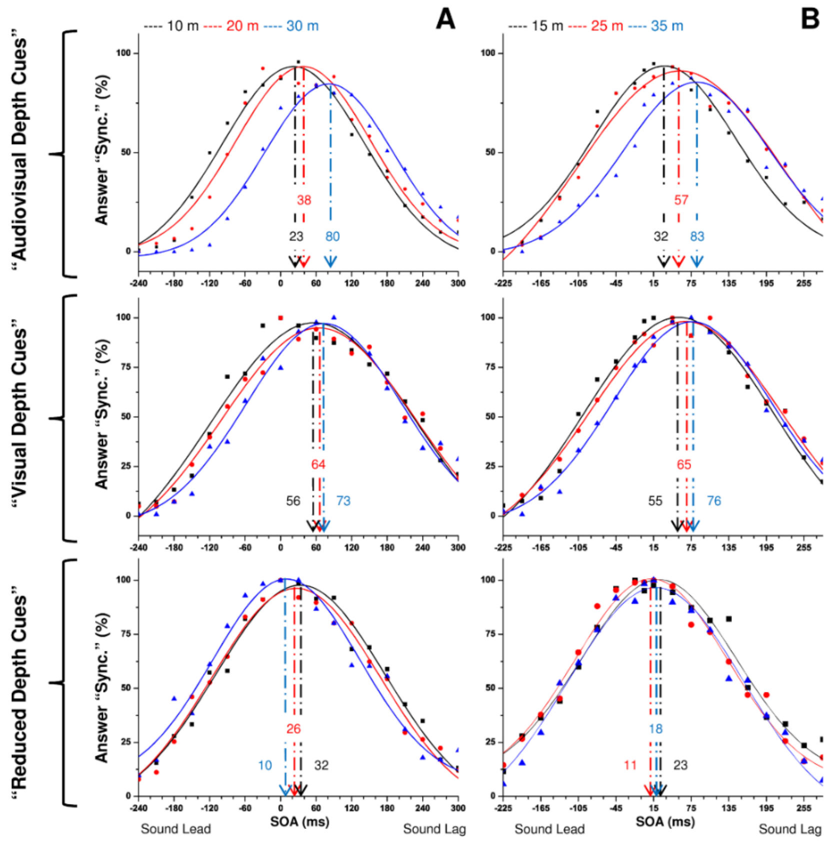


Figure 37. Proportion of "synchronized" answers as a function of the SOA for a data pool of distances of 10, 15, 20, 25, 30 and 35 m in the three experimental conditions. Each proportion was calculated using a pooled data of 160 answers from the 4 participants. Panel A shows the answer distributions for distances of 10, 20, and 30 meters; panel B shows the answer distributions for distances of 15, 25, and 35 meters. A fit of a Gaussian function was performed in order to get the PSS and WTI values for each distance of stimulation. An arrow indicates the PSS for each stimulus distance.

In both the "audiovisual depth cues" and "visual depth cues" conditions, the peak of the Gaussian curve progressively moves towards a higher sound delay as the stimulus distance increases. This increment is generally lower in the "visual depth cues" condition, where there is a difference of only 20ms between the lowest (in the 15 m presentation) and the highest (in the 35 m presentation) PSS, while in the pooled data of the "audiovisual depth cues" condition this difference is of 60ms (with the lowest PSS at the 10 m presentation and the highest at 35 m). However, in the "reduced depth cues" condition the peak of the Gaussian curve hardly moves from one distance to another (especially in the odd group of distances) and when it moves, it does so in the direction of a decrement in sound delay. This was the only condition where several PSSs with a value close to zero (see distances 25 and 30 m) were found. A One-way ANOVA shows significant differences between conditions regarding the PSS (F (2, 15) = 11.9, p < .01), and the Scheffé Post-hoc tests revealed that these differences are significant when we compare the PSSs in the "reduced depth cues" condition with the PSSs in the "visual depth cues" condition (p < .01) and with the PSSs in the "audiovisual depth cues" condition (p < .05).

Figure 38, plots the increment in the PSS as a function of the increment of the stimulus distance regarding the first distance of presentation for each condition. Here, we can compare the way the PSS changes across conditions with a model for internal compensation of the slower propagation velocity of sound. A linear function was fitted to the PSSs obtained in the three conditions. A good adjustment (r^2 = .94, F (1,4) = 84.8, p < .01) on the fitting of the function *y = 2.6x* to the results in the "audiovisual depth cues" condition was obtained. In the "visual depth" condition the best linear function was *y = 0.9x*, with a good fit (r^2 = .93; F (1,4) = 63.84, p < .01). Similarly, a linear function was fitted to the PSSs obtained in the "reduced depth cues" condition (*y = -.75x*), but this time with only with a rough adjustment (r^2 = .5, F (1,4) = 5.77, p < .1).

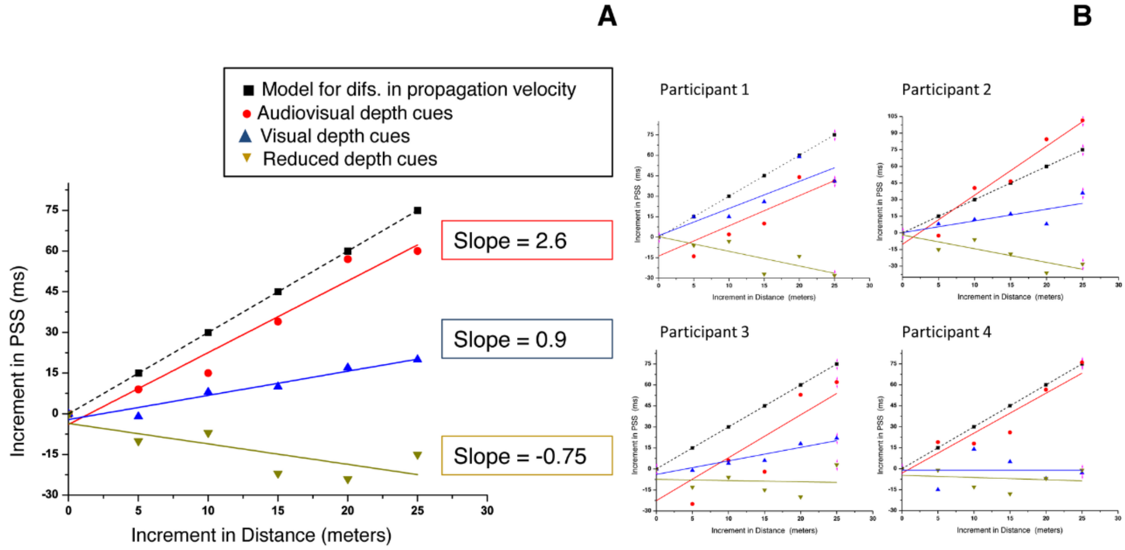


Figure 38. Increment in the PSS as a function of increment in the stimulus distance, regarding the first distance of presentation. Black squares correspond to the theoretical values predicted by a mechanism that compensates for differences in propagation velocity. Red dots are the PSS found for each distance in the "AV depth cues" condition. Blue triangles are the PSS found for each distance in the "visual depth cues" condition and orange triangles are the PSS found for each distance in the "reduced depth cues" condition. A fit of a linear function was performed in each group of data. **Panel A** shows the graph for the pooled data, and **Panel B** show the same graph for the individual data.

Figure 37 and **Figure 38** clearly show that the conditions "audiovisual depth cues" and "visual depth cues" present different tendencies when compared with those from the condition "reduced depth cues". While the PSS from the "audiovisual depth cues" and "visual depth cues" condition increases with distance, the PSS from the "reduced depth cues" condition seems to slightly decrease with distance. In fact, correlation tests show that the PSS in the "audiovisual depth cues" condition is positively and significantly correlated with distance (r_{sp} = 1, p < .001) and the same is true for the PSS in the "visual depth cues" condition (r_{sp} = .943, p < .01). On the other hand, PSS in the "reduced depth cues" condition is just marginally correlated with distance (r_{sp} = -.77, p < .1), and in the opposite sense: higher PSSs are associated with lower distances.

Discussion: Our study aimed at verifying if a perceptual mechanism that compensates for audio-visual asynchronies as a function of the presentation distance exists, and if so, how can it be expressed in a PCM capable of guiding IVE developers in the implementation of accurate aural-visual synchronization. To do so, we presented audiovisual biological stimuli at several distances and with a wide range of stimulus onset asynchronies in an IVE. Crucially for this study, we manipulated the amount of distance cues in three experimental conditions: audiovisual depth cues, visual depth cues and reduced depth cues.

Results from the first two experimental conditions revealed a systematic shift of the point of subjective simultaneity in the direction of greater audio lags with greater distance. Therefore, results support the existence of an internal compensation mechanism for varying stimuli delays with distance. Interestingly, compensatory evidences were not found in the "reduced depth cues" condition. Indeed, that condition was so impoverished regarding distance perception cues that only visual angular velocity – a relatively poor distance cue – was available. Therefore, the internal compensation mechanism appears to be dependent on the amount and quality of depth cues simulated in the IVE. This interpretation is further supported by the fact that in both conditions where evidence for such mechanism was found, the steepness of the function was not the same. In the "audiovisual depth cues" condition there was a steeper function, closer to that expected from the actual physical audio-visual delays. We conclude that the proposed compensation mechanism might not work in an all-or-nothing way, but there might exist intermediate levels of compensation. However, at this stage we are not able to provide an exact account on what is guiding a compensatory pattern of response in some conditions. It could be argued that such a mechanism emerged due to the enhanced realism of the stimuli, or due to the causal relation between visual and auditory stimuli. In any case, we can assume that both factors contributed to a greater perceived unity between the multisensory signals. Future studies should focus on the relative effects and weights of different cues, different stimuli, and different settings and, with this in mind, a critical review of previous studies in this field might reveal that data obtained so far was not necessarily contradictory, but mostly the result of different experimental setups.

Despite all these limitations, some hypotheses on what is guiding compensation can be further discussed, considering our results. First of all, assuming that our stimuli were perceived as co-localized in the "audiovisual depth cues" condition, co-localization of auditory and visual stimuli seems to be an important factor to exhibit compensation for the relatively slow speed of sound. However, co-localization does not seem mandatory for compensation of sound propagation velocity. Alais and Carlile have found evidence of compensation for sound propagation velocity by providing auditory depth cues while keeping the visual stimuli at a fixed distance [106]. In their study the ratio of direct-to-reverberant energy was used as an auditory depth cue, but the visual stimulus was fixed at 57 cm from the observer and primarily used as a reference point in time. Furthermore, their results show that the compensation effect relies on the robustness of the auditory depth cues. Thus, the authors concluded that reliable auditory depth cues together with a task-relevant situation are sufficient in order to activate compensation for the sound propagation velocity. These results shift the focus from co-localization to specific depth cues, when we try to uncover the reason for having compensation in some experimental conditions. Our results agree with the idea that a powerful auditory depth cue is necessary in order to get evidence for sound propagation velocity compensation: only when we add the binaural recordings of sound steps do we get clear evidence for such compensation.

Our work was carefully designed taking into account several problems pointed out in previous studies. More specifically, we avoided the use of simple and stationary stimuli with poor distance cues. We also used audiovisual stimuli with a causal relation between them, as well as familiar stimuli. By using a PLW and real step sounds, we provided the participant with a type of stimulus that occurs in everyday life. Indeed, the finding of compensation for sound propagation velocity with such stimulus might be evidence that, apart from relying in some depth cues, this mechanism also uses knowledge from previous experience. In real-life situations, we are always exposed to a certain audio delay. Despite this delay being highly variable, a long time of exposure to it could lead to some kind of temporal recalibration. This approach is in accordance with the work of Heron and collaborators where, after a brief phase of exposure to the natural sound lag of a distant audiovisual stimulus, participants shifted lag expectations [121]. Moreover, the work of Virsu and collaborators has shown that simultaneity constancy can be learned in natural interactions with the environment and without explicit feedback [122]. Effects of simultaneity constancy appear to be long-lasting and modality specific. In addition, studies comparing the perception of synchrony in adults and infants have found that the thresholds for asynchrony detection are modified as we get older [123], further showing that synchrony perception is affected by our history of exposure to certain audio delays.

For IVEs developers, these findings highlight the importance of including naturalistic sound delays in highly immersive audiovisual virtual scenes. As we can see from our data, IVEs users can judge audiovisual synchrony and are sensitive to small variations in the audio-visual temporal relation. Thus, the rule of delaying the sound output about 3ms per meter, depending on the distance of the audiovisual stimulus, should be applied. Moreover, and by looking at the size of the WTIs from our pool of participants, we can understand how important it is to accurately measure end-to-end delays of IVEs and to struggle to keep these delays at a minimum level. On average WTIs for our participants where about 125ms (σ = 48ms), which means that this is the range of asynchronies around the PSS at which the participant will still respond with at least 75% of "synchronous" answer. It is important to note though, that this range is evenly distributed around the PSS and therefore is moveable, meaning that due to vision-first bias and effects of distance on the PSS this WTI shifts towards being increasingly tolerable to sound lag values and increasingly intolerant to vision lag values.

In conclusion, from this work we can derive three important design guidelines for IVEs developers:

1. End-to-end delays of visual and auditory output should be measured and controlled in a way that a visual and an auditory stimulus modeled to be perceived synchronously, should not lag more than 62.5ms (half of the average WTIs value);

2. On an audiovisual synchronic scene, the vision-first bias should be taken into consideration and sound should never precede image;

3. The natural occurring sound lag should be taken into account and modeled when developing the IVEs. The rule for a buffer responsible for defining the sound delay should be:

Equation 10. $y = y_0 - k + (3.4x)$*, where* y_0 *is the measured difference between the audio and the visual end-to-end delay of a specific IVEs, k is a constant delay inserted to compensate for this difference, and x is the distance-to-user of the simulated audiovisual event. All variables are in milliseconds.*

In this work, we have been looking into how temporal disparity affects perception of audiovisual stimuli in IVEs. In the next experiment, we will look into the role of spatial disparity and its effects on audiovisual perception in IVEs.

3.2. Audiovisual Unity Assumption

3.2.1. On the Role of Spatial Coincidence

The design and equipment constraints of most IVEs require output of the different modality stimulus to be transmitted through non-co-localized means. For example, in a projection-based audiovisual IVE, the visual stimulus is projected on a screen while the auditory stimulus is conveyed through headphones or a set of speakers. This is particularly relevant if we consider that often the goal of an IVE is to present a *unitary* audiovisual stimulus at a given depth and position in a 3D space. As we saw in **Section 3.1**, this goal can be achieved by using convincing simulations of both visual and auditory depth cues. Binocular and pictorial depth cues allow users to perceive visual stimuli beyond the projection screen; while binaural cues and room acoustics allow users to perceive sound in space. Thus, in a perfect audiovisual virtual system, the projection screen fades away and gives room to perfectly defined objects in distance and the headphones wear off as the sound is perceived to be outputted from an external source in space and not from the apparatus in your head. The problem though is that, as we saw in **Section 2.6,** perfect virtualizing systems do not yet exist.

While congruent audiovisual depth cues are simpler to manage in systems without user tracking, consider again the problem of end-to-end latency in interactive IVEs. Both visual and auditory end-to-end latency will result not just in temporal mismatch but also in physical displacement of the virtualized audiovisual object relative to the users' head. Moreover, if the end-to-end delay is different for the visual and the audio output, users will also experience physical displacement between the relative position of the visual and the auditory stimulus. Thus, it should be quite relevant for IVEs developers to know and to be able to accurately measure the just noticeable amount of physical disparity between the visual and the auditory stimuli – i.e., the displacement value above which an audiovisual stimulus stops being perceived as such and starts to be perceived as two separated stimuli, one visual and one auditory. This value would give developers information about how manageable their end-to-end latency is (and if it is interfering with the understanding of the audiovisual scene); while a method to accurately measure this value, would give developers a technique to test mitigation measures.

One of the main perceptual effects developers rely on in order to deal with audio-visual spatial incongruence is simply trusting on a well-known human perceptual phenomenon called the *Ventriloquism Effect (VE)* [124], [125]. The VE, also known as *capture effect*, happens when sound is perceived as coming from the location of a visual stimulus, despite being spatially displaced. The term is based on the trick used to give the impression of a dummy speaking, when in reality is its operator emitting the sound.

The VE is constantly present when we are watching a movie either at home or at the cinema, as sound is coming not from the actors' location on the screen but from speakers located elsewhere. In real setting environments, the limits of VE were quantified to be around 20° azimuth, depending on the direction of the spatial disparity – the limit is smaller for displacements in the frontal area (around the 0° azimuth) and higher for displacements in medial areas (around the 45° azimuth) [126].

Although we can trust in VE to deal with most of the problem of audio-visual displacement in many audiovisual applications, we should not rely on VE if we want to design an IVE where precise auditory location is required. Kytö and collaborators [127], described an AR application where stereoscopic visual and binaural auditory stimuli where used in order to signalize *points of interest* (POI) in a navigation system. Because these POIs would have to be precise in the tens of angular units and because they needed to use acoustic POIs to direct user attention to portions of the virtual environment outside the visual FoV, they needed to know the limits of VE in order to effectively use these acoustic POIs (i.e., the POI sound should not be captured by anything being visual displayed in the scene as that could trigger an error in the users navigation). In order to accurately measure the limits of VE, Kytö and colleagues used an Oculus Rift AR system and Pure Data for audio rendering using HRTFs from the CIPIC database [128] for auralizing the auditory stimuli. They presented the participants with multiple possible combinations of auditory-visual stimulus displacement (see **Figure 39**) and asked participants to respond to the question: "Is the sound (audio POI) coming from the speech bubble (visual POI)?"

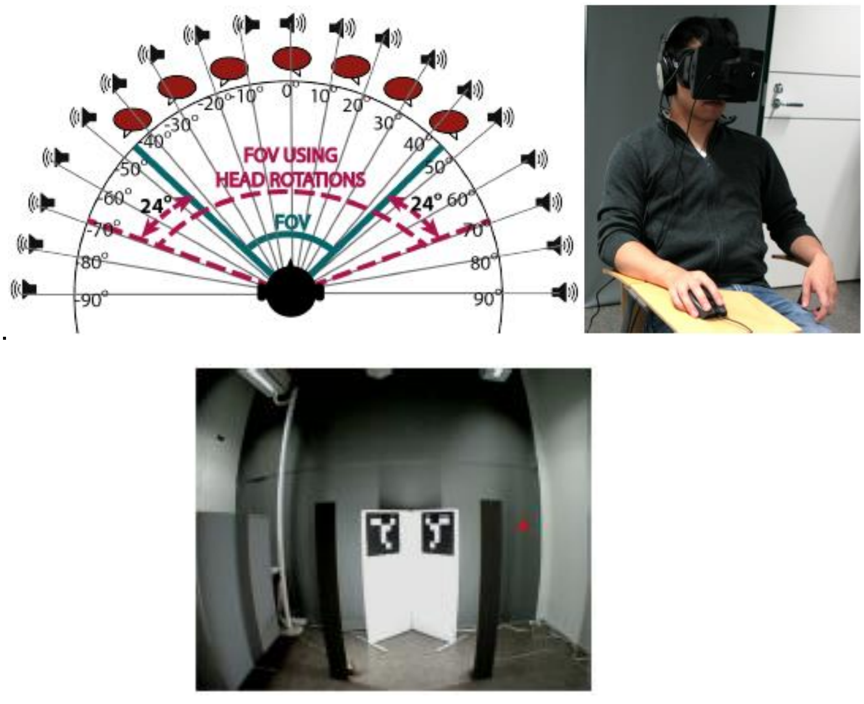


Figure 39. Set-up of Kytö et al. experiment. Top figure, shows all the possible combinations (in angular position in azimuth) of the audio-visual stimulus presented to the participant [127].

Using the Just Noticeable Difference (JND, see **Section 3.1.1**) as the limit at which 75% of the participant's answers reported separation between visual and auditory stimulus, Kytö and colleagues found values between 32° and 45°, depending on the visual angle (mean JND was 38°). Similar to results from [126], participants needed further sound displacement to judge audio-visual separation for visual presentations around the 45° azimuth, than for visual presentation around the 0° azimuth location. Thus, the main guideline to extract from [127] is that, if we want to clearly spatially separate the visual from the auditory stimulus in an IVE, we should do so by keeping a spatial disparity greater than 32°-45° depending on visual direction. Of course, this interesting result can be interpreted from another perspective, which is: if the spatial disparity between visual and auditory stimulus is below 15° (average value of the 25% level of answers "stimulus are separated"), they are most likely to be perceived as an unitary audiovisual stimulus.

As we saw with the work of Kytö and colleagues [127] AR navigation systems are one of the applications requiring precise spatial and temporal fidelity (see also [129] for an AR application on acoustic navigation). Other applications, such as interactive virtual concert halls or multi avatar VR environments, would greatly benefit of a clear knowledge of the spatial limits of audiovisual UA.

3.2.2. Assessing the Spatial Limits of Unity Assumption

Thus, in this experiment we wanted to device a way of accurately measure the limits of UA in an IVE. Differently from [127], we were interested in measuring audiovisual unity assumption in a set where multiple visual stimulus compete for the correct attribution of the auditory source. This is closer to applications such as virtual concert halls or multi avatar environments, where the user should correctly bind the sound to each agent in the virtual scene (e.g., a music sound to an instrument in an acoustic setting, or step-sounds to avatar in a simulated room).

Our main goal with this experiment is to accurately define absolute spatial thresholds for a correct unity judgment between an auditory stimulus and its correspondent visual stimulus in a conflicting task. Furthermore, we also wanted to explore if those limits are dependent or independent from stimuli distance.

Experiment 1 – Method

Participants: Four participants, aged 19-30 years old, underwent visual and auditory standard screening tests and had normal hearing and normal vision. All were voluntary participants, and all gave written informed consent to participate in the study.

Stimuli and Materials: The experimental tasks were performed in the set-up described in **Section 3.1.2**. Additional materials used in the implementation of this experiment can be found at: https://github.com/Carlos-CCG/Thesis2019/tree/master/ExpUnityAssumption

A) Auditory Stimuli.

The auditory stimuli were step sounds from the database of controlled recordings from the College of Charlston [Marcell2000]. They correspond to the sound of a male walking over a wooden floor and taking two steps at exactly the same velocity as the visual stimuli. These sounds were auralized as free field by a MATLAB routine with HRTFs from the CPIC database[28] [128]. With this custom MATLAB routine, a binaural sound would be generated through the application of the correct ILD and ITD cues (based on a geometric model describing sound source position in the 3D space) and the nearest HRTF would be choose from the CPIC database in order to perform the last step of head-related frequency modulation. A set of 20 different position sounds were generated and used in this experiment (from azimuth -45° to azimuth 45°, in steps of 5° plus two steps auralized at 3° and -3° azimuth).

B) Visual Stimuli.

Visual stimuli were two PLWs (similar to the ones described in **Section 3.1.2**) located at 20 meters in a front-parallel plane to the observer. The PLWs were back-to-back and equidistant from the center of the projection screen. Throughout the duration of a trial, each PLW would take two steps, but with the translational component removed.

C) Auditory and Visual Stimuli Relation.

The audio-visual scene consisted in the presentation of the two PLWs at a given separation degree and the output of sound co-located and synchronized (a sound delay of 68 ms was added, based on the 20 m of distance of the visual stimuli) with just one of the PLWs. Thus, the angular distance between the PLWs would randomly vary from trial to trial on a range of 45° horizontally, in steps of 5° plus a stimulus with just 3° of separation (**Figure 40**).

[28] Available at https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/

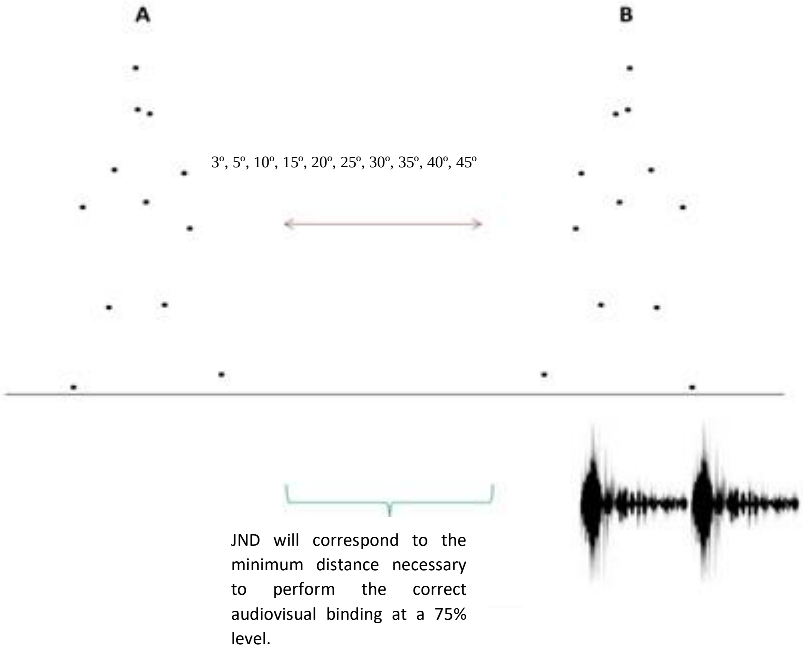

A
B
3º, 5º, 10º, 15º, 20º, 25º, 30º, 35º, 40º, 45º
JND will correspond to the minimum distance necessary to perform the correct audiovisual binding at a 75% level.

Figure 40. A graphical representation of the auditory and visual stimuli relation in this experiment.

There were a total of 20 different audiovisual stimuli (10 angular distances between PLWs x 2 sound bindings, left PLW or Right PLW). Each stimulus was presented 40 times, with a total duration of 4.5s (3s of stimulus presentation + 1.5s of inter-stimulus interval). Thus, the total experiment took 1 hour to be complete for each participant (four sessions of 25 min. each).

Procedure: Participants were first briefed on the purpose of the experiment and then proceeded to seat in a position aligned with the center of a 2.80 x 2.10m projection screen. The participant's head rested on a chin-rest to prevent lateral movements. The task participants were asked to perform was: "Please pay attention to the following audiovisual scenario. You will see to PLWs at a certain lateral distance from each other and taking two steps. You will hear the sounds of two steps synchronized and co-localized with one of the PLWs. The sound will be co-localized with either the PLW from the left or the PLW from the right, please if you think the sound is from the left PLW press the left button of the mouse, if you think the sound is from the right PLW press the right button of the mouse."

Results: Figure 41, shows the response distribution for each participant. A Maximum Likelihood Estimation (MLE) function was fitted to the data of each participant, using RStudio and a package called 'quickpsy' [130]. The MLE is based on a Logistic function of the type:

Equation 11. $y = \frac{L}{1+\, e^{-\beta\,(x-x_0)}}$ *, where L is the curve's maximum value, β is the curve's slope, and x_0 is the curve's inflexion point.*

We got a satisfactory fit of the MLE function for three out of the four participants. Participant 3 had the poorest fit level, as it failed to give responses with a high level of confidence even at the presentations where PLWs were more separated.

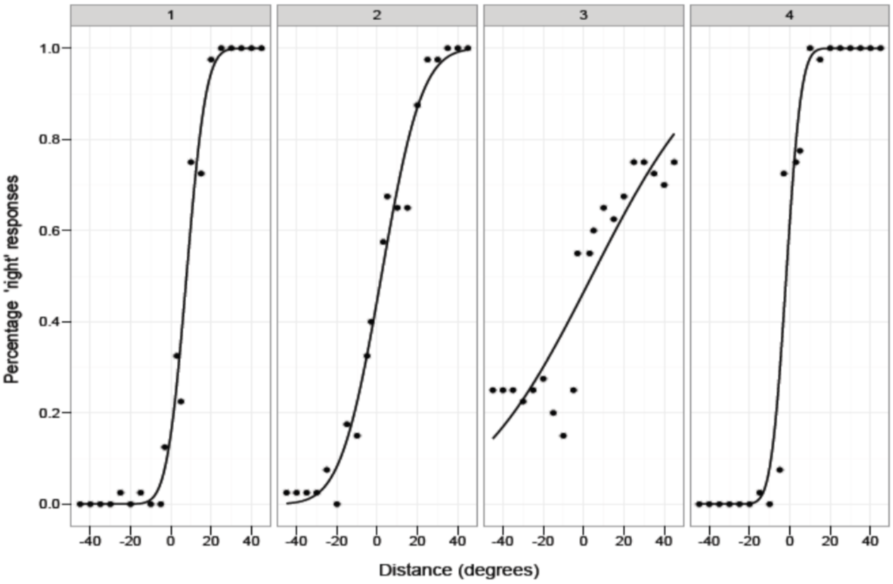


Figure 41. Individual data for the UA Judgment Task. The graphs show percentages of responses "right" as a function of the PLWs' degree of separation. Negative values of separation mean presentations where sound was co-localized with the PLW in the left.

The MLE estimation outputs the *Point of Subjective Equality* or PSE (the curve's inflexion point), the β (curve's slope), and allow us to extract the JND (the point where participant's distinguish correctly the audio-visual binding within a 75% success rate). **Figure 42**, shows the pooled data of participant 1, 2, and 4 (participant 3 was excluded due to poor adjustment of the model to the data).

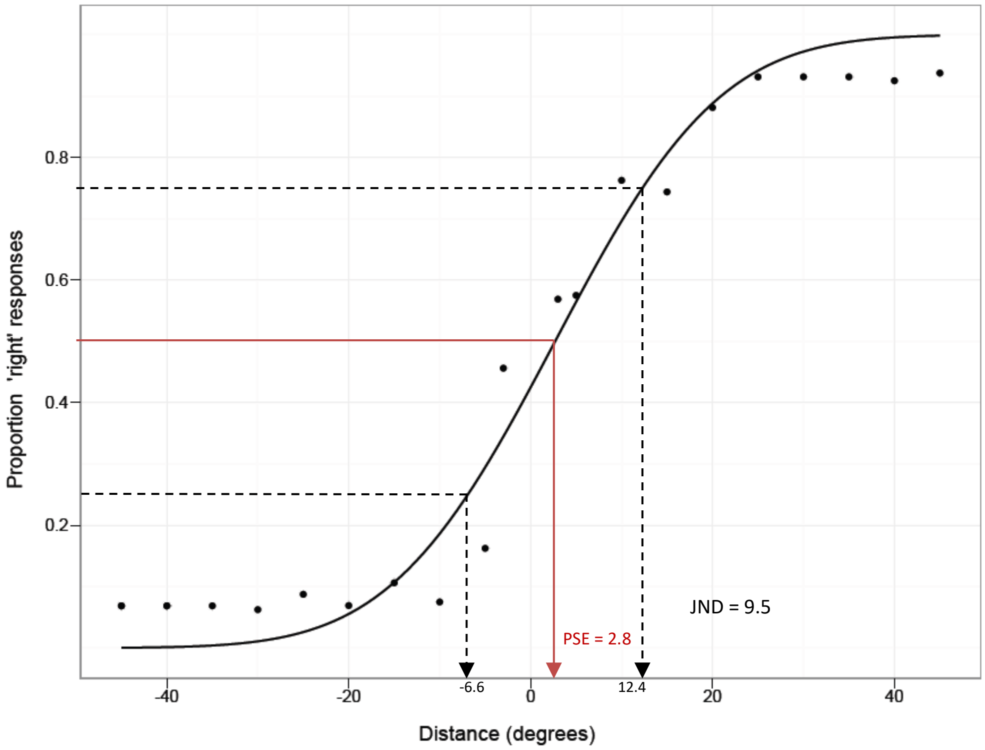


Figure 42. Pooled data for the UA Judgment Task. The graphs show percentages of responses “right” as a function of the PLWs’ degree of separation.

We used a method to find the JND in Temporal Order Judgments tasks described in [44]. According to these authors, we can calculate the JND of a cumulative psychometric function by dividing by two the difference between the independent variable value at the 25% and at the 75% point. In our case the separation degree between PLWs that elicit 25% of responses “right” was -6.609° (negative value means that sound was coming from the PLW in the left), and the separation degree that elicit 75% of responses “right” was 12.408°. Therefore, our JND was [12.408 – (-6.609)]/2 = 9.508, meaning that at 20 meters participants needed a minimum separation of 9.5 degrees between conflicting visual stimuli to correctly binding a sound to the visual form of its source.

Discussion: The UA judgement task developed in this experiment, appears to successfully capture the threshold for correct bindings of auditory and visual stimulus in an audiovisual scene with multiple conflicting visual stimuli. Three of four participants manage to perform the task and inter-participant performance was similar. Interestingly enough, when we pooled the data we found a smaller JND than those founded in both [127] and [126]. The JND in our task (9.5°) actually halved the JND found by Hairston et al. [126] for frontal stimuli presentation (20°) and it is less than half the value founded by Kytö et al. [127] (32°). These differences might be due to the different nature of the experimental tasks.

In fact, while Kytö and Hairston were concerned with giving an account of absolute VE thresholds between one single visual and auditory stimulus, our task was design in order to give an account of threshold for a correct UA in a conflict task where sound should be correctly bind with one of two possible visual stimuli. It is important to note that the presence of a conflicting visual stimulus apparently helped participants to perform the correct audiovisual binding, when compared with previously published tasks of absolute UA judgment between a single visual and auditory stimulus.

Thus, our task expands the guidelines regarding audiovisual UA presented by Kytö on [127], by assessing UA in a virtual environment with competing visual stimuli. However, and before summarizing the guidelines for IVEs developers that one can extract from this work, we wanted to further explore a possible effect of presentation distance on the JND in UA judgment tasks.

Hairston and colleagues [126] showed that variability in the localization of the stimuli were strongly correlated with biasing effects in a VE task. In other words, the JND increased as the audio and visual stimuli were presented in locations were its absolute position was harder to judge (i.e., positions distant from the participants' midline point and towards peripheral locations). It is known that performance in distance estimation is negatively correlated with the stimulus presentation distance [92], thus in a second experiment with UA judgment tasks we wanted to explore the role of stimuli distance of presentation. Finally, we also wanted to assess participants' performance in a purely auditory task in order to understand how good of a predictor of performance in an UA judgment task participant's auditory acuity can be.

Experiment 2 – Method

Participants: Three participants, aged 22-28 years old, underwent visual and auditory standard screening tests and had normal hearing and normal vision. All were voluntary participants, and all gave written informed consent to participate in the study.

Stimuli and Materials: The experimental set-up was similar to the one used in experiment 1.

A) Auditory Stimuli.

Similar to experiment 1, the auditory stimuli were step sounds from the database of controlled recordings from the College of Charlston [120]. They corresponded to the sound of a male walking over a wooden floor and taking two steps at exactly the same velocity as the visual stimuli. These sounds were auralized as free field by a MATLAB routine with HRTFs from the CPIC database[29] [128]. With this custom MATLAB routine, a binaural sound would be generated through the application of the correct ILD and ITD cues (based on a geometric model describing sound source position in the 3D space) and the nearest HRTF would be choose from the CPIC database in order to perform the last step of frequency head-related frequency modulation. An attenuation for different presentation distances was included based on the *inverse square law* [131], a standard equation for acoustic attenuation from atmospheric absorption [132]:

Equation 12. $A_a = -20\, log_{10} \left[\frac{L_d}{I_0}\right]$ *, where A_a is total attenuation in dBs, d is distance, L_d is path length, and I_0 is sound intensity at d = 0.*

A set of 18 different angular sounds positions were generated and used in this experiment (from azimuth -40° to azimuth 40°, in steps of 5° plus two steps auralized at 3° and -3° azimuth). These sounds were generated for two presentation distances (20 and 10 meters). An additional set of 18 sounds were auralized at 10 meters but at the same metric variation as the sounds at 20 meters (therefore at higher angular positions).

B) Visual Stimuli.

Similarly to experiment 1, the visual stimuli were two PLWs in a front-parallel plane to the observer. The PLWs were back-to-back and equidistant from the center of the projection screen. Throughout the duration of a trial, each PLW would take two steps, but with the translational component removed. PLWs were presented at two different distances (20 and 10 meters), plus a third condition were PLWs were presented at 10 meters but with the same metric disparity of PLWs at 20 meters (therefore at higher angular positions).

C) Auditory and Visual Stimuli Relation.

Overall, there were two conditions regarding type of stimulus – an audiovisual and an auditory condition.

In the auditory condition sound was presented at two distances (10 and 20 meters). In the "20 meters" condition, sound position varied on a range of 40° horizontally, in steps of 5° plus a stimulus with just 3° of separation. In the condition "10 meters with the same angular disparity", the metric

[29] Available at https://www.ece.ucdavis.edu/cipic/spatial-sound/hrtf-data/

horizontal distance to the center of projection of the sound stimuli position were shortened to equalize the angular disparity of the "20 meters" condition. Finally, in the condition "10 meters", the horizontal distance to the center of projection consisted in the angular disparities obtained by keeping the metric distance of the "20 meters" condition constant and approaching the stimuli 10 meters, therefore we had presentations at: 5, 10, 18, 29, 38, 47, 55, 64, and 72° of disparity (see **Figure 43** for an example).

In audiovisual conditions, the scene at each trial consisted in the presentation of the two PLWs at a given separation degree and the output of sound co-located and synchronized (a sound delay of 68 ms was added, based on the 20 m of distance of the visual stimuli) with just one of the PLWs. Thus, the angular distance between the PLWs would randomly vary from trial to trial.

In the condition "20 meters", where stimuli were presented at 20 meters from the participants, the distance between PLWs varied on a range of 40° horizontally, in steps of 5° plus a stimulus with just 3° of separation. In the condition "10 meters with the same angular disparity", the metric distance between PLWs was smaller to equalize the angular disparity of the "20 meters" condition. Finally, in the condition "10 meters", the distance between PLWs consisted in the angular disparities obtained by keeping the metric distance of the "20 meters" condition constant and approaching the stimuli 10 meters, therefore we had presentations at: 5, 10, 18, 29, 38, 47, 55, 64, and 72° of disparity (see **Figure 43** for an example).

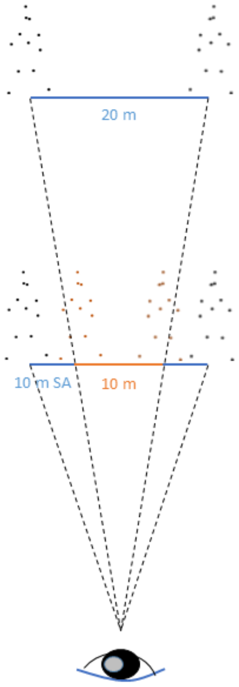


Figure 43. A geometric demonstration of the difference between angular and metric disparity.

There were a total of 18 different stimuli (9 angular distances between PLWs x 2 sound bindings, left PLW or Right PLW) presented 20 times, with a total duration of 4.5s (3s of stimulus presentation + 1.5s of inter-stimulus interval). There were 6 different experimental conditions, thus the total experiment took 1 hour and 42 min. to be complete for each participant (divided in 12 sessions of about 13 min.).

Procedure: Participants were first briefed on the purpose of the experiment and then proceeded to seat in a position aligned with the center of a 2.80 x 2.10m projection screen. The participant's head rested on a chin-rest to prevent lateral movements. The task participants were asked to perform was: "Please pay attention to the following presentations. You will be presented with two different conditions regarding type of stimulus. In some experimental blocks you will see two PLWs at a certain lateral distance from each other and giving two steps. You will hear the sounds of two steps synchronized and co-localized with one of the PLWs. The sound will be co-localized with either the PLW from the left or the PLW from the right, please if you think the sound is from the left PLW press the left button of the mouse, if you think the sound is from the right PLW press the right button of the mouse. The PLWs will be presented at different distances from each other and this will vary within experimental blocks. Finally, in some experimental blocks you will only hear the step sounds. In these trials you should use the mouse to indicate the sound position relatively to the screen midpoint, by pressing the left button if you think sound is coming from the left, or by pressing the right button if you think sound is coming from the right."

Results: In **Figure 44**, we present a MLE function fitted to a data pool of the three participants, for each one of the conditions in experiment 2.

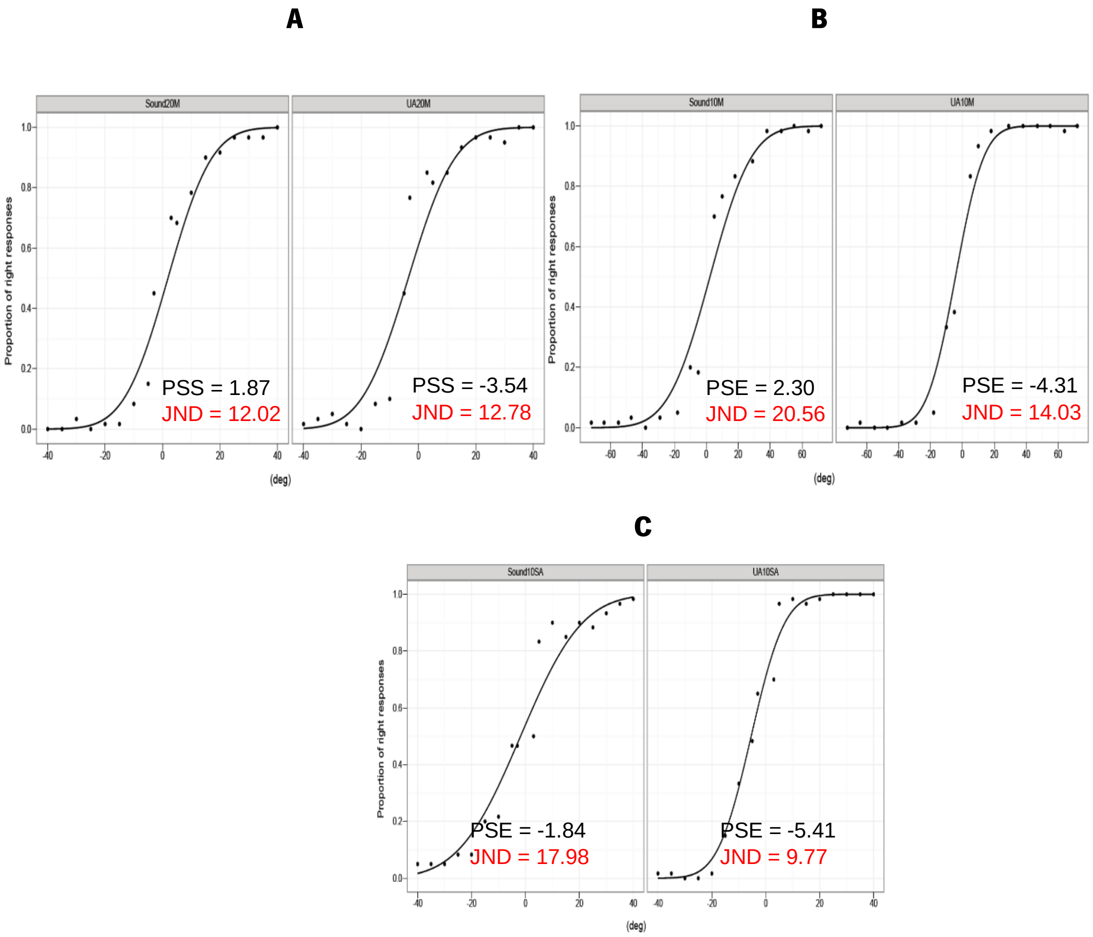


Figure 44. Pooled data for the "sound only" condition and for the audiovisual conditions at 20m (A), 10m (B), and 10m same angular disparity (C). The graphs show percentages of responses "right" as a function of the PLWs' degree of separation.

We can observe on **Figure 44** that there is not a clear relation between JNDs and distance of presentation. Values vary between 9.77° and 14.03° for presentations at 10 meters, and 12.78° for presentation at 20 meters (replication of experiment 1, which obtained a JND of 9.51°).

Nevertheless, one interesting result is the lower JNDs for audiovisual presentations (however, see 20 meters condition), meaning that spatial coincidence between two signals of different modalities might help to the precision on this task. This is a result that might indicate a benefit of multisensory presentation for localization tasks, and wich might be further investigated in future research.

Discussion: In both experiments, we found smaller JNDs, in the audio-visual tasks, than those found in both [126] and [127]. As we referred above, this can be an indication that having conflict stimuli of one modality (e.g. visual) competing for the binding of the second modality (e.g. auditory) helps performing a correct binding in a UA judgment task. Moreover, it also appears to enhance localization performance when compared with unimodal auditory tasks, at least at shorter distances from the observer ("sound only" conditions in experiment 2).

The two experiments taken together allow us to produce additional guidelines for IVEs developers regarding spatial simulation of audiovisual stimuli. [127] advised developers to use a maximum angular disparity of 15° if they want to convey the perception of one audiovisual POI in an AR scene, or conversely to use angular disparities between 32° and 45° if the goal is to convey the perception of two separate stimuli, one visual and one auditory. From our experiments, we can conclude that in order to avoid wrongfully UA attributions, IVEs developers should:

1. Avoid audio-visual spatial mismatches that render the audio at a position closer than 9.5° - 12° azimuth from a conflicting visual stimulus;

2. When possible, co-localized visual stimuli should be used in order to improve spatial discriminability of auditory stimuli;

In the next experiments, we will explore some phenomena of unimodal perception in IVEs and we will continue to demonstrate examples of psychophysics application as a method to study and improve perception in these environments.

3.3. Perception of Visual Distances in Virtual Environments

3.3.1. On the Need to Access Basic Perceptual Phenomenon to Validate Simulators for Studying Complex Perception-Action Behavior

One of the most valuable applications of IVEs is as a platform for controlled experimentation or training. Using IVEs removes the risks inherently associated to ecological scenarios and brings the advantages of controllability to environments that are, perception wise, close enough to the real-world scenario that we may say they have *ecological validity*. If IVEs were capable of producing an exact replication of the real-world event with action-perception correspondence, we could refer to the IVEs as in a state of *stimulus fidelity* [133]. However, both technological limitations of the immersive system and the user's prior knowledge of physically being in a place other than the one conveyed by the simulation, make impossible for contemporary immersive virtual systems to reach such a state. As a result, IVEs users often notice differences between the simulation and the simulated environment, also called *non-identities*.

Accordingly, IVEs developers should always perform two types of control activity when designing a new virtual environment for any interactive application. The first control activity consists in comparing real-world scenarios and IVEs, to assess if the participants are noticing non-identities (assessment of the IVEs stimulus fidelity). The second control activity consists in assessing if the detected non-identities impair performances, i.e., whether the performances in an IVE are similar to the participant's performances in real-world scenarios (assessment of the IVE *action fidelity*, also known as *face validity* [133], [134]). Psychophysical experimentations offer a quantifiable evaluation of human performances in perception-action related tasks, thus enabling comparisons with user's performances in the real world and consequently making possible to access the user's identification of IVEs non-identities and the IVEs action fidelity [135]–[137]. Furthermore, another way of looking into the action fidelity of an IVE is to try a replication of known real-world perceptual phenomena in an IVE.

These control activities become even more important as today IVEs are being widely employed as training platforms as well as in a wide variety of psychological experimentation such as on visual [138] and auditory [139], [140] perception and action, spatial cognition [141], and even on social interaction [142], [143]. In the realm of controlled experimentation, some IVEs are being developed to study perception-action behaviors as complex as crossing the road (**Figure 45**). Before performing control experimentation to study complex perception-action behaviors, an IVE developer must have means to evaluate the participants' perceptual mechanisms wich are crucial to the experimental goals. For instance, in the case of using pedestrian simulators like the ones in **Figure 45** to study road crossing behavior – where the pedestrian has to accurately judge the distance of an approaching vehicle, calculate the time it would take him to get to the other side of the road, and judge the risk of a collision – the experimenter has to guarantee that the IVE is conveying an *accurate sense of distance perception*. In fact, accurate distance perception as since long being regarded as critical for interaction and navigation in IVEs [136], [144].

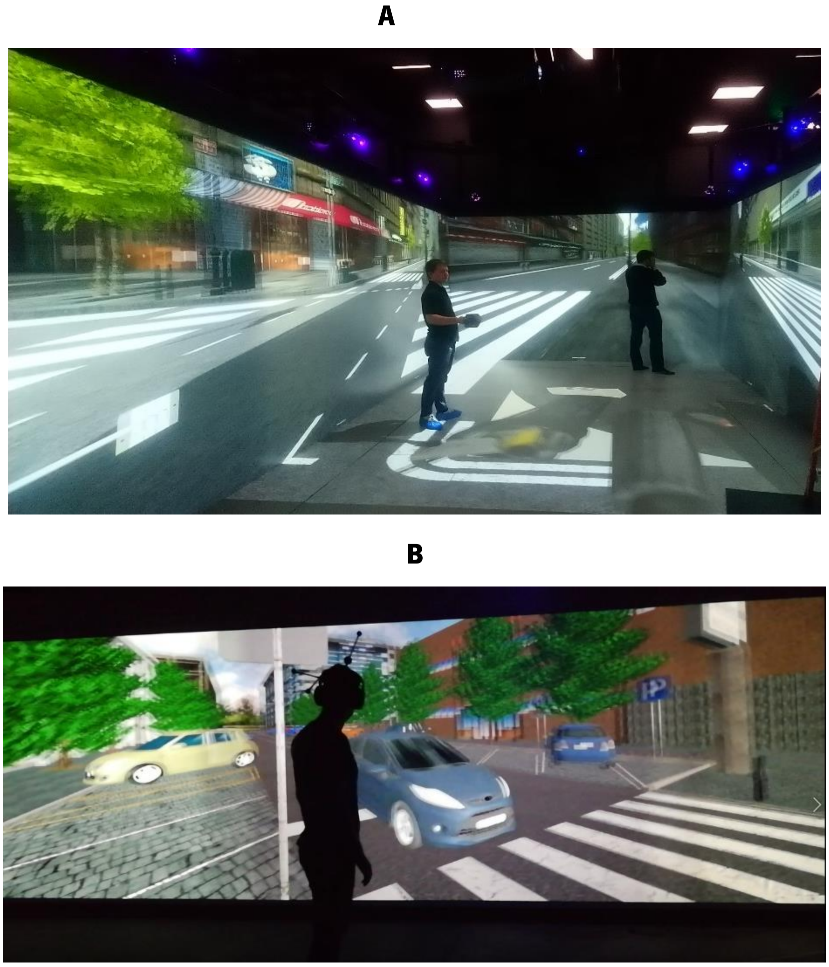


Figure 45. Two pedestrian simulators, that resort to projection-based visual IVEs, currently being used in transportation research. **Panel A**, is the University of Leeds Highly Immersive Kinematic Experimental Research (HIKER) simulator, Image from the University of Leeds. **Panel B**, is the pedestrian simulator located at the Center for Computer Graphics, Guimarães and developed in the context of the ANalysis of PEdestrians Behaviour (ANPEB) project.

In the present study, we used a psychophysical experiment that we thought is ideal to evaluate stimulus fidelity and action fidelity of an IVE used in a distance perception task. We used a psychophysical task named *Frontal Matching Distance Task* (FMDT) [145] in order to evaluate participant's distance perception. Comparable measurements of distance perception in an IVE and a real-world scenario are particularly appropriate to access both the stimulus and the action fidelity of an IVE. It is known that humans have a tendency to underestimate egocentric distances in a FMDT [145]–[147]. Experiments using a range of distances that goes from 5 meters up to 25 meters revealed a pattern of response showing that egocentric distance is perceived as compressed, but at a nearly constant ratio [145]. Gathering data from a FMDT in a real-world scenario and in an IVE with different conditions varying the level of simulation realism allows to check for perception of non-identities in the simulated environment and, furthermore, will provide new insight into the importance of photorealism in IVEs (see [148], for an account of the debate). Moreover, looking at patterns of progressively underestimation with increasing distance will show if participants are sensitive to some perceptual effects similar to those that affect observers when in interaction with real-world environments.

3.3.2. Measuring Distance Perception: A Comparison Between Real-world and Simulated Scenarios

Methodology

The FMDT was first thought as an egocentric distance perception task by Li and collaborators [145] (although similar methodologies were already presented in [149] and [150]). Results on egocentric distance estimation tasks have been revealing that egocentric distance is normally perceived fairly linearly, however far from accurate. The typical pattern of response, on this type of task, is an underestimation of distance by a nearly constant ratio [145]. The FMDT is a particularly suitable task to assess stimuli fidelity of an IVE, because it only requires the presentation of two marks with a given distance between them on a frontal-parallel to the observer plane. Thus, transposition of the FMDT from the real world to an IVE is fairly easy and the data obtained in the latter environment is directly comparable with the data obtained in a real-world scenario.

Participants: Five participants, aged 24-29 years old, underwent visual standard screening tests and had normal, or corrected to normal, vision. All were voluntary university students or researchers and all gave written informed consent to participate in the study. All the five participants were naïve as to the purpose of the experiment.

Stimuli and Materials:

1. **Real-World Scenario (RWS)**: The experimental task in the RWS condition was performed in a grassy open-field located in Campus de Azurém at the University of Minho (see **Figure 46**). Two collaborators were used as marks and the participant was standing facing one of them. Two sisal yarns were laid and stretch on the ground to form perpendicular lines. The participant was instructed to walk on top of the sisal yarn in order to ensure their perpendicularity to the marks' plane. The sisal yarns had different lengths so that its end could not be used as a distance cue. The distances between the two collaborators used as marks and between the participant and the facing mark were measured using a Bosch™ DLR130k laser measure.

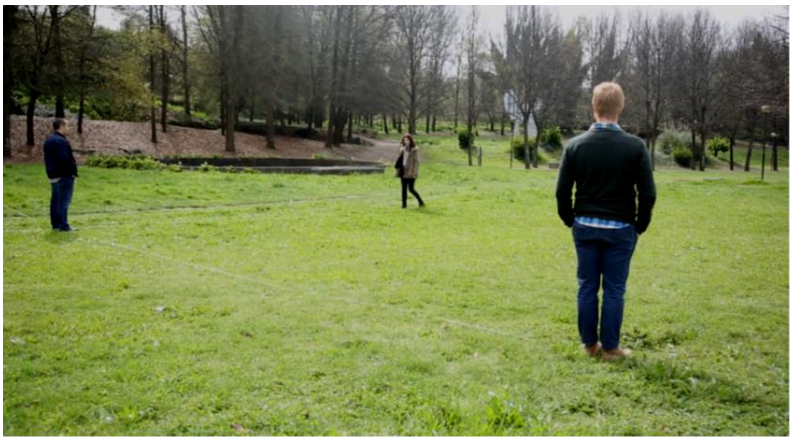


Figure 46. FMDT in the RWS condition. Participant performing the FMDT in an outdoor location at the University of Minho.

2. **Immersive Virtual Environment (IVE):** The experimental task in the IVE conditions was performed in a darkened room at the Centro de Computação Gráfica in the Campus of Azurém, University of Minho. For the stimuli presentation we used a cluster of PCs with NVDIA® Quadro K5000 graphic boards and the virtual environment was designed in the Blender software with projection conFigd and computed through the Blender VR software. Two 3chip DLP projectors Christie Mirage S+4K and one DS+6K-M SXGA+ DLP Christie projector with a resolution of 1400x1050 pixels and a frame rate of 120Hz were used for the stereoscopic projection (using active stereoscopy) of the visual stimuli. The area of projection was a power-wall like screen of 2.80 m high and 9 m wide. A Vicon® motion capture system composed of six cameras was used to track the participants head position and update the projection accordingly (see **Figure 47**). Two experimental conditions were presented, differing from each other in the level of photorealism of the virtual environment.

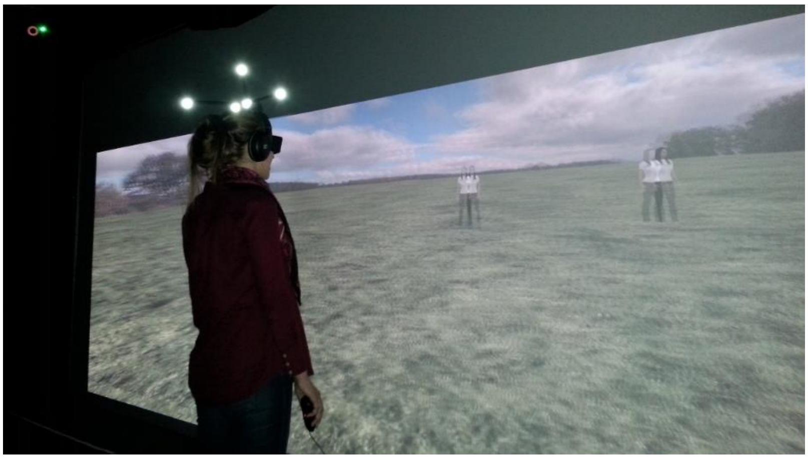


Figure 47. FMDT in the IVE conditions. Participant performing the FMDT in the IVEPH condition. The structure with retroreflective markers on the participant's head is used in order to track participant's head position and rotation.

The *"Photorealistic Immersive Virtual Environment (IVEPH)"* condition presented a similar environment to the one in the RWS condition, with two avatars as marks located in a grass field with trees at the back (far enough to preclude their use as a relative size cues). Using a Photo Research® Photometer/Colorimeter Model PR 655 that outputs values of color space and luminance, the colors of the grass and the sky in the IEPH condition were adjusted to match the colors of this elements in the RWS condition. A model of atmospheric scattering was use in order to simulate aerial perspective (see **Figure 48** – A).

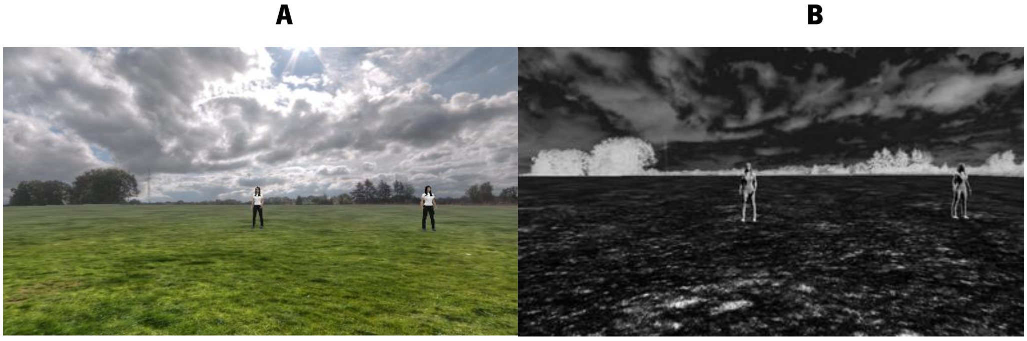


Figure 48. IVEs rendition of the FMDT. (A) Depiction of the virtual environment used in the IVEPH condition (B) Depiction of the virtual environment used in the IVENPH condition.

The *"Non-photorealistic Immersive Environment (IVENPH)"* presented the same environment as in the IVEPH, however the colors of all the presented textures suffered a monochromatic negative transformation (see **Figure 48** – B).

Procedure: In each trial, the participant was presented with two marks (two persons in the RWS condition and two avatars in the IVEPH and IVENPH conditions) in a given fronto-parallel plane (L-shape arrangement, see **Figure 49**.The participant's task was to position him(her)self at a distance from mark 1 equal to the distance separating marks 1 and 2.

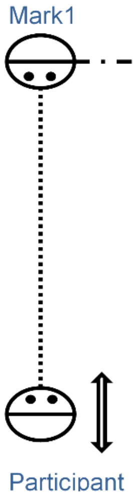


Figure 49. Schematic representation of the FMDT. Participants had to move towards or recede from Mark1 in order to equalize their egocentric distance to Mark 1 with the distance between Marks.

The participant's initial distance to mark 1 and the distances between mark 1 and mark 2 varied across trials. Three initial participant-mark 1 distances (8, 14, and 20 meters) and nine mark 1-mark 2 distances were used (5 to 21 meters by step of 2 m). The participants performed 2 sessions of 27 trials (3 participant's initial distances x 9 distances between marks) in each environment condition, all presented in a random order.

At the beginning of each experimental session, the following instructions were given: "You are going to participate in a study on distance perception. In this task you will have to judge the distance between two persons, the ones located in front of you, and adapt your distance to the person directly in front of you so that it matches the distance between the two persons. You can move towards or recede from the person in front of you as many times as you wish, however you cannot move laterally." In the RWS, participants moved forwards and backward by walking and indicated to the researcher when they reached the appropriate location and the researcher then measured the participant-mark 1 distance. In the IVEs conditions participants moved forwards and backward, by moving themselves in the space available in front of the projection screen and by using respectively the right and left button of a mouse. They pressed the wheel button when they were in the correct position and their location was saved by the computer before starting the next trial.

The virtual FMDT was implemented in Blender (release 2.78c) [BlenderFoundation2018], an open-source 3D computer graphics software, written in C, C++, and Python, with an integrated game engine called Blender Game.

Commands and Logs

Using Blender's game logic we implemented the controls for starting the experiment, controlling the movement of the participant's position, and confirm the participant's final position in each trial (**Figure 50**).

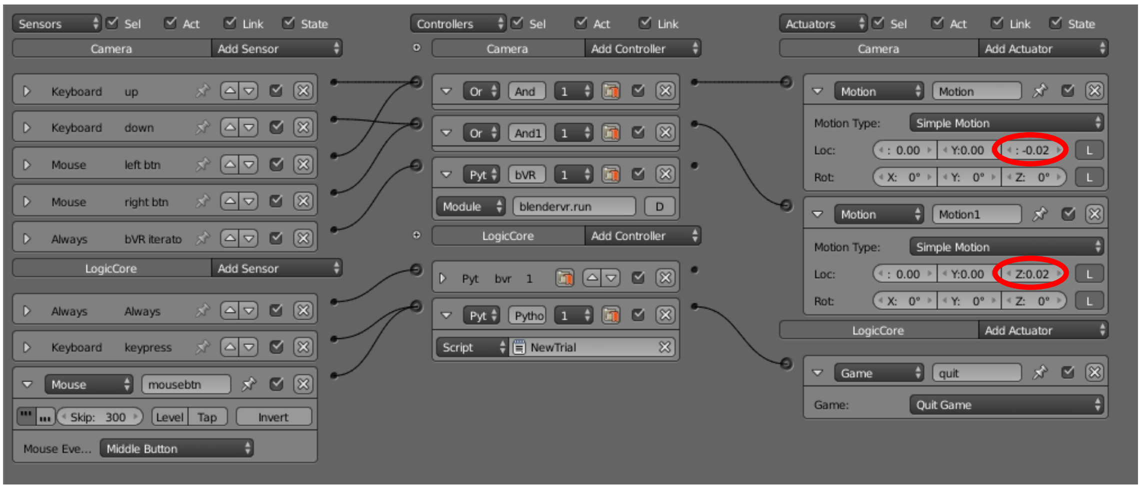


Figure 50. A simplified game logic for the virtual FMDT.

UP and DOWN keys or the LEFT or RIGHT mouse buttons would, respectively, approach or recede participant's camera position in relation to Mark 1. Thus, the participant could combine physically approaching to Mark 1 (by walking towards the projection screen) with the use of the game commands to walk the last meters in order to perform the FMDT. The camera's height, rotation, and distance to origin is defined through motion capture data coming from the Vicon® system, but a camera offset locked in the Z axis (use as depth) is possible using the above mentioned commands for approach or recede (see red circles in **Figure 50**). The final distance-to-mark1, takes into account tracked position and camera offset in the Z axis.

A Python script runs every time the participant confirms the final trial answer by pressing the middle button on the mouse. The script logs the participant final position and defines the next trial, as follows:

```
INITIALIZE an array with all possible positions between marks [10 positions]
INITIALIZE an array with all possible initial positions participant-to-mark [3 positions]

IF Position Matrix is not created:
    INITIALIZE a Matrix with all possible position combinations
    INITIALIZE an array with a SHUFFLE of all possible combinations
```

```
ELSE Save trial number, distance between marks, and participant position

POP one trial from the Positions Matrix
ASSIGN a position from the position Matrix to each mark PLUS a new participant initial distance
```

Avatars

The avatars consisted in two identical female characters in a normal standing position. The avatars were generated using the *Make Human* software (version 1.1.1) [151] and later rigged in Blender using a 30 bones armature, allowing to change their final pose. This project can be downloaded at: https://github.com/Carlos-CCG/Thesis2019/tree/master/ExpDist

Results: The individual analyses showed similar response patterns across participants in the FMDT (i.e., the condition that showed the highest and lowest mean error was the same for all participants). Therefore, individual data were pooled and mean values were analyzed using regression analysis. **Figure 51** shows the fit of linear functions to the data pooled as a function of the experimental condition. Data distribution, grouped by condition, conformed well to the linear fittings (mean Pearson's r=.99). Due to a general underestimation of egocentric distance, participants placed themselves too far from mark 1 when matching the mark1-mark2 distance, in all conditions. However, there was a significant effect of the condition on the FMDT mean error (Wilks' Lambda = .38, F(2,134)= 110.8, p<.01), due to the fact that the mean error in the RWS condition (2.3m), was lower than the mean error in the IVEPH (5 m), and IVENPH (6 m) condition (both p<.01). The two latter conditions were also different (p<.01).

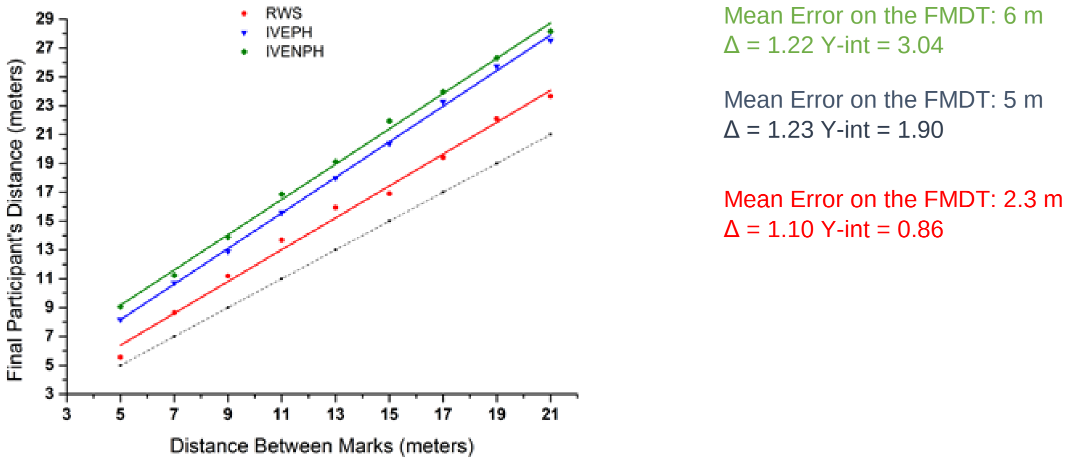


Figure 51. Final participant's distance as a function of the distance between marks. The black dashed line represents veridical performance (i.e. no error in the FMDT). Red dots are FMDT mean results for the RWS condition. Blue triangles are FMDT mean results for the IVEPH condition. Green squares are FMDT mean results for the IENPH condition.

Discussion: The first conclusion we can draw from this data is that having a photorealistic IVE with perspective and stereoscopic depth cues is not enough to elicit a real-world performance in a distance judgment task. Nevertheless, a photorealistic IVE minimizes the discrepancy between performance in simulated environment and real world, when compared with a non-photorealistic IVE. The importance of photorealism has been the focus of a heated debate in the past 10 years, and studies have provided inconsistent results. Scerbo and Dawson [150] argued that high-fidelity simulators do not always lead to better performance and, in some cases, can interfere with the performance. In the same sense, Sanchez-Vives and Slater [146] called attention for the lack of evidence supporting realism as an important contributor for the feeling of presence (see also [152] and [153]). However, in 2009 Slater reviewed his position when his team showed that greater visual realism increases the feeling of presence [148] (comparison between IVEs generated using real-time recursive ray tracing and IVES generated through ray casting). Likewise, Yu and collaborators [154] showed that the feeling of presence in an IVE was increased by the use of global illumination, as well as dynamic shadows and reflections (see also [155]). Following Yu and collaborators, our results are also supporting the idea that realistic illumination favors a realistic performance in a perception-action task. Our results revealed however a stronger support for photorealistic IVEs than those reported by Thompson and collaborators [153], and we may hypothesize that the emergence of results in favor of high-fidelity IVEs in the recent years is linked to the better graphic potential of today's systems. We consider that the difference that we got between performance in the IVEPH condition and performance in the real-world condition is still liable to be reduced through the increment of our IVE's immersion capacity. This could be achieved, for example, by adding a floor projection channel that would provide the participant's with additional cues that could be used in the distance judgment tasks.

The second conclusion that we can draw form our results is that, despite the differences in the absolute mean error, the pattern of responses was quite similar in all conditions. Slopes of linear regressions are positive and slightly higher than one, which means that the error increases with the distance roughly the same way in all conditions. This finding indicates that, despite failing to elicit a real-world response regarding the absolute value of performance (*absolute validity*), our IVEs are capable of eliciting a like real-world pattern of responses (*relative validity*). An increase of underestimation when increasing the distance is a common pattern of results in egocentric distance judgment tasks [134] [144], which we obtained in both of our IVEs conditions.

We still do not know if an exact replication of the real-world performance (i.e. stimulus fidelity) in our IVE would be possible in a near future. Nevertheless, our experimental set-up remains an appropriated tool to evaluate the gap between performance in IVE and in real-world scenario in distance judgement tasks. It is implemented in a cross-platform software and is simple enough to configure for the accurate projection in different IVEs technological supports (both projection and display-based). A summary of the guidelines provided by this experiment are:

1. The FMDT is an appropriated task to measure the validity of IVEs requiring the performance of distance judgments by the users.

2. Real-world distance judgement follows a linear function with a slope higher than one, meaning that the natural occurring distance judgement error increases with the distance of the visual stimulus. In our study the PCM for real-world distance judgments is:

Equation 13. $y = 0.86 + (1.1x)$, *where x is the physical distance of the visual stimulus in meters.*

3. Photorealism is minimizes the discrepancy between performance in simulated environment and real world, probably by reducing IVEs user's detection of non-identities.

In this experiment, we showed how psychophysical studies can be used to evaluate an IVE in terms of stimulus and action fidelity. This approach can be quite useful to validate IVEs to be used in interactive applications as different as training and simulation for highly specialized operators, high-end visualization (e.g. scientific visualization, prototyping), teleoperation, and human behavior studies and social interactions. Furthermore, we showed that an IVE can have action fidelity (i.e. elicit a similar to real-world pattern of response and real-world perceptual phenomena) without achieving a state of full stimulus fidelity (i.e. perfect perceptual simulation of the real-world), difficult to achieve with the technological limitations of today's immersive virtual systems.

3.4. Spatial Discriminability of Auditory Virtual Stimuli

3.4.1. On the Rising of Commercial Applications for 3D Sound

The recent growth of immersive technology in the consumer market is mainly due to new technological advancements in visual displays. Nonetheless, both researchers and manufacturers have already acknowledged that for successful immersive experiences, it is also important to create an appropriate and congruent immersive listening environment [157]–[159]. Audio output devices, particularly wearable ones, will play a major role in the transition between commercial visual immersive systems to commercial audiovisual immersive systems [159].

The Earphones and Headphones industry has been steadily growing following the emergence of new technological advancements – as noise canceling and wireless technology – and new applications – like the incorporation of 3D sound in virtual reality systems. As applications requiring spatialized sound make their way into the market, new methods to determine listeners' performance will be in high demand. These assessments are of particular interest for IVEs developers that are looking for the best audio devices to support auditory stimulation. The integration of spatial sound in IVEs has been positively correlated with the feeling of presence [160] and the IVEs industry is already aware of the benefits that one can gather when more effort is focused on sound rendering (see, for instance, the collaboration between Oculus Rift and RealSpaceTM 3D audio).

As we described in **Section 2.1.6**, the most widespread method for auralization and acoustic simulation takes into account the listener's anatomy – head, pinnae, and ear canal shape – and simulates its effect on the sound wave [161]. The listener's anatomy affects mainly the inter-aural time and inter-aural level differences (ITD and ILD respectively), which are the main static cues for sound location [162]. Thus, we can simulate a given position of the sound source in azimuth and elevation, by filtering an anechoic sound through a function that shapes each channel output giving it the accurate ITD and ILD for that position in space. These functions are called Head Related Transfer Functions (HRTFs). Auralization using HRTFs seems to be an appropriate solution for commercial applications, particularly the ones using databases of non-individualized HRTFs (captured using Head and Torso simulators). Studies have shown that listeners can locate non-individualized HRTF-based sounds [162] and that short training sessions improves significantly the localization performances [163].

In this study, we present a method that allowed us to find out if listener's performance on auditory location tasks using non-individualized HRTFs is dependent on the type of audio devices used. This question is particularly interesting when we compare headphones and in-earphones, because the former devices allow individualized pinnae and ear-canal modulation over the non-individualized HRTFs, while the latter devices do not.

3.4.2. Measuring Auditory Location of Virtual Stimuli as a Function of the Audio Device

Methodology

Participants: 16 participants with no previous experience in laboratory controlled auditory location tasks. All participants had normal hearing, measured by standard audiometric tests. None showed inter-aural sensitivity differences above 5dB HL.

Stimuli and Material: A three second duration anechoic Pink Noise, auralized using HRTFs taken from the MIT database [164]. We present 18 different source positions in the horizontal plane (i.e., elevation 0°), with azimuth ranging from front to right in steps of 6°, from azimuth -6° to azimuth 96°. All sounds were auralized as free field presented at 1 meter from the listener. The sound output intensity was measured and matched for both audio output devices, using a Brüel & Kjær type 4128C head and torso simulator and a PULSE™ acoustic analyzer platform.

Two conditions were developed, regarding audio output device (headphone VS in-earphone) in the experimental phase (intra-subjects). The headphones used in this experiment were the Sennheiser HD 650 (**Figure 52** – A) and the in-ear phones were the Etymotic ER-4B Micro Pro (**Figure 52** – B). Both audio devices show a linear response on the range of 100-1000 Hz. Furthermore, the participants were divided in two groups regarding the audio output device used in training phase (inter-subject). Meaning that eight participants performed a training session with the headphones and the other eight performed a training session with the in-ear phones.

Figure 52. Audio output devices used in the experiment. (A) Headphones – Sennheiser HD 650; (B) In-earphones – Etymotic ER-4B Micro Pro.

Procedure: The overall experiment consisted of three phases:

(1) *Pre-training phase* where all stimuli were randomly presented (with four repetitions each). After each stimulus presentation its localization was estimated in a touch-screen (see **Figure 53**, panel A);

(2) *Training phase* where for five minutes participants could freely listen to five stimulus correctly positioned in the answer interface (see **Figure 53**, panel B). At the end of this time, participants listened each one of the five trained sounds and should click on the correct rectangle. Correct feedback was given at the end of each trial and this phase would end when participants reached an 80% correct answer level of performance;

(3) *Post-training phase*, where participants repeated the same procedure as in the pre-training phase.

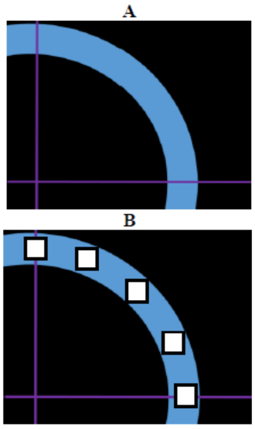


Figure 53. Answer interface. *Panel A* – Participants were required to estimate the sound position in azimuth along the purple arch. *Panel B* – 5 positions of the trained stimuli. The answers were collected in a touchscreen, using a touchscreen stylus in order to increase precision.

The experiment was programed using the XML-based BioMoSe language (see **Example 2**) to define each scene and the presentation of the auditory stimuli was controlled by *VR Juggler* [165]. Logging of the participant's responses was performed by establishing a TCP/IP protocol that would save the pixel coordinate of each tap in the touchscreen with scene indexation, in a central data repository. The pixel coordinate were then converted in degrees using the formula: $Degree = arctan\ (y/x)$, and setting the origin (0;0 – i.e. the point where the two pink lines cross in **Figure 53)** at the pixel coordinate (38;158) – counting from the superior-left corner of the touchscreen.

Additional materials used in the implementation of this experiment can be found at: https://github.com/Carlos-CCG/Thesis2019/tree/master/ExpAuditoryLocation

Results: Table 7, shows the absolute mean error in degrees azimuth for each audio device used in each experimental phase.

Table 7. Performance by condition and experimental phase

N = 16	Data Grouped by Device Used in Experimental Phase	
	Azimuth Pre-Training	*Azimuth Post-Training*
Headphones **Abs. Mean Error**	16.77° (SD=4.79)	13.88° (SD=4.62)
In-earphones **Abs. Mean Error**	18.83° (SD=6.74)	17.97° (SD=8.04)

The absolute mean error is lower on the Headphones condition, for both the Pre-training and the Post-training sessions. Paired sample t-test revealed significant differences between listening devices for the absolute mean error in the Post-training session (t (15) = -2.513, p<.05). A difference of 4.02° in the post-training results, corresponds to a sound displacement of approximately 7 cm, at 1 meter from the listener. **Figure 54** presents the absolute mean errors distribution as a function of the stimuli position, for both conditions.

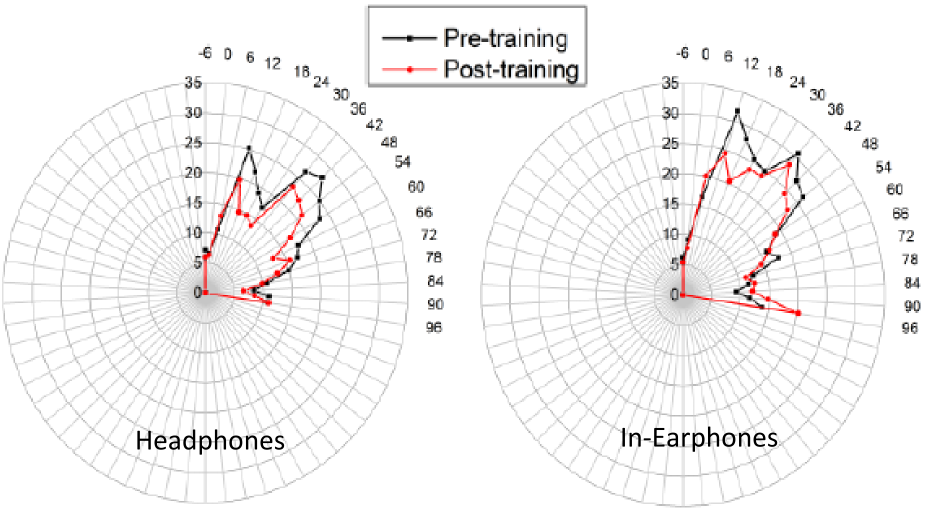


Figure 54. Polar graphics with the absolute mean error as a function of the stimuli's position.

As we can see from **Figure 54**, the localization errors are higher in intermediate azimuths and lower on the ear plane and on frontal regions. This pattern of response is present with both equipment; however, there are globally lower errors in the headphones condition and that is even more clearly observed in the extreme presentations (ear plane and frontal regions).

In a second analysis, we grouped the participants by audio output device used during training sessions. In doing this, we wanted to understand how congruency regarding devices used on training and experimental sessions might affect performance on auditory location.

Table 8. Data grouped by training listening device.

N = 8	**Data Grouped by Training Listening Device - Headphones**	
	Azimuth Pre-Training	*Azimuth Post-Training*
Headphones Abs. Mean Error	17.24° (SD=4.97)	13.84° (SD=5.24)
In-earphones Abs. Mean Error	17.64° (SD=5.61)	20.13° (SD=10.33)
N = 8	**Data Grouped by Training Listening Device – In-Earphones**	
	Azimuth Pre-Training	*Azimuth Post-Training*
Headphones Abs. Mean Error	16.36° (SD=4.94)	13.93° (SD=4.25)
In-earphones Abs. Mean Error	19.85° (SD=7.97)	15.81° (SD=4.85)

From **Table 8** we can see that keeping congruency (grey cells) between listening devices used during training and experimental phases, gives rise to generally lower absolute mean errors of sound localization in the post-training phase. Incongruence between training and experimental session listening device disrupted completely the benefits of training in the case of participants that used in-earphones in experimental phases. A mean decrement in performance of about 2.5° is observed for these participants, from pre to post-training session (also the mean value presents more variability). Nevertheless, incongruence did not prevent learning and better performance in post-training sessions for participants that used headphones in experimental phases.

Figure 55 presents the distribution of the mean error as a function of the stimuli position, for the congruent sessions (same audio output device in training and experimental sessions). In **Figure 55**, positive errors indicate misjudgments in sound location towards the ear plane, while negative errors indicate misjudgments of sound location towards the frontal plane (azimuth 0°).

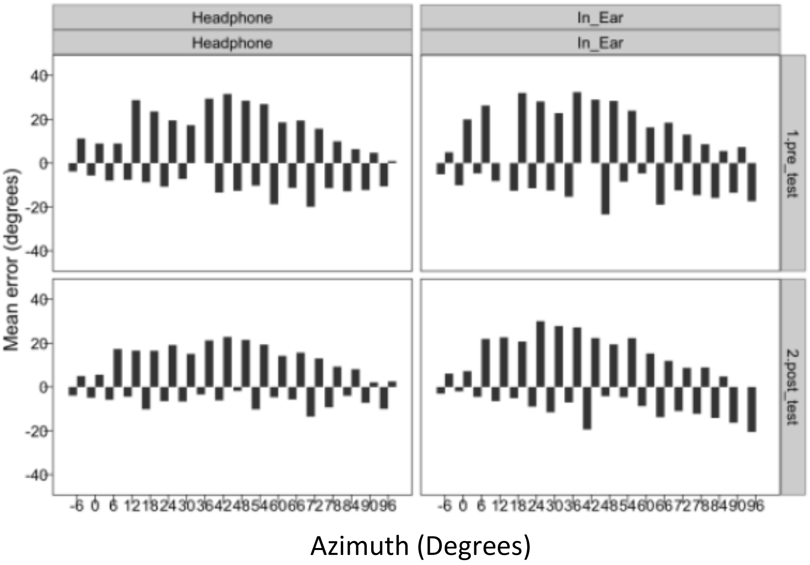


Figure 55. Mean error distribution and direction as a function of the stimuli position, for the congruent sessions. Positive errors indicate misjudgments in sound location towards the ear plane, negative errors indicate misjudgments of sound location towards the frontal plane.

Interestingly, it is possible to observe that positive errors are predominant, meaning that when misjudging location participants are prone to locate the stimulus as closer to the ear plane.

Finally, as headphones are more permeable to external noise when compared with in-earphones, we conducted a test to verify if the results obtained in silent conditions would hold in conditions with added environmental noise. Thus, we replicated this experimental protocol for eight new participants in a set-up in which the environmental noise reached the 56 dB(A) SPL. In these environmental conditions, participants had an absolute mean error of 18.52° azimuth for the pre-training session, and an absolute mean error of 14.86° azimuth for the post-training session. These results differ on an average of 1.16°, when compared with results of congruent sessions using headphones.

Discussion: We presented a valuable method to assess listener's spatial perception and evaluate performance between two audio devices. In the comparison between these particular models, headphones appeared to be the best solution for presentation of auralized sound and we should further investigate the benefits of using large housing with open back headphones. The fact that large housing headphones may allow individualized *pinnae* and ear-canal modulation over the non-individualized HRTFs, might be an important factor in the final performance outcome.

Nevertheless, we can reduce the differences between devices if short training sessions are included and the same audio output device is used between training and test. In-earphones can benefit greatly of maintaining congruency between experimental and training phases. Future work should exhaustively compare between several types of audio output devices and should investigate how performance is affected by the introduction of binaural room acoustic cues.

From this work, we can conclude additional guidelines for IVEs development:

1. Choosing audio output devices should be guided by the evaluation of user's perception in controlled conditions;

2. Open-cascade headphones, which allow individualized pinnae and ear-canal modulation over the non-individualized HRTFs, are preferable over listening devices that have in-ear sound output.

3. When using non-individualized spatial cues, a short training session is advisable and can result in considerable improvements in location tasks. However, for the training improvements to be transferrable from training sessions to generic use situations, one should maintain the displaying or audio output device.

3.5. Conclusion

In this chapter, we argued that the empirical approach, which was prominent at the dawn of HCI, could be particularly useful in this era of renewed and widespread use of IVEs. In fact, the same empirical approach that resulted in the development, and ultimately adoption, of certain UIs that are still part of today's personal computers will play the same role in order to define which IVE systems will be further developed and eventually be part of the long desired IVE as an UI. Thus, the main value of this section is in the experimental tools and protocols here described and their pertinence for the development of PCMs. Accordingly, its success should be measured by how IVE researchers and developers adopt these experimental methodologies.

As we pointed out at the end of **Section 1.3**, there are two types of user's cognitive models capable of providing insights into design and interaction problems. In **Part II**, we have been providing practical demonstrations on the role of PCMs, thus in **Part III** we will the same for DCMs.

Part III

DCMs and Safety-Critical Interactive Computing Systems

4. THEORETHICAL BACKGROUND ON MEDICAL DEVICES ICSs AND DCMs

In this Chapter we will introduce the reader to safety-critical ICSs, which will be our use case to develop new DCMs capable of detecting instances of potential use-error and guiding the definition of design guidelines capable of mitigating them. Due to the consequences of use errors when using such safety-critical ICSs, these devices constitute the perfect use cases to illustrate the usefulness of DCMs, an approach specially tailored to be applied in early design stages or during the investigative process towards understanding use-error in a particular UI. Medical devices are a category of safety-critical ICSs, especially prone to use-error with drastic consequences. Moreover, the area of medical devices is an emergent area in terms of new software and increasingly complex UIs. These devices stopped being manipulated by experts only in controlled (however sometimes harsh) environments such as the emergency room and began to be used also by patients with any kind of background and in a wide range of contexts of use.

Next, we will better expose what are safety-critical interfaces and what are the HCI challenges in medical devices.

4.1. What are Safety-Critical Interfaces?

Complex safety-critical systems are defined in [166] as:

> "… *a system whose safety cannot be shown solely by test, whose logic is difficult to comprehend without the aid of analytical tools, and that might directly or indirectly contribute to put human lives at risk, damage the environment, or cause big economical losses.*"

As Marco Bozzano points out on his book *Design and safety Assessment on Critical Systems* [167], this definition is peculiar because it chose to merge two different concepts that could be defined separately, *complexity* and *criticality*. Nevertheless, the motivation to presenting them together has to do with the fact that these concepts often go hand in hand with safety-critical systems, where there is a steady trend to increasing complexity as these systems become gradually more prevalent. There are many traditional applications of safety-critical systems in areas such as aviation and aircraft flight control, military and aerospace, and energy systems such as the nuclear industry. Moreover, the complexity of current communication and information systems is contributing to the development of new areas of application of safety-critical systems in industries such as the automotive and healthcare.

In the scope of this thesis, we are mainly concerned with the aspects related to the UI of safety-critical medical devices. Thus, complexity is relevant as long as it influences the quality of the interaction, as it often does. The increasing complexity of the software layers of a safety-critical system – that are "invisible" to the user – often result in medical devices that have a great number of functions, and where distinction between states is sometimes hard to comprehend and test [167].

The history of safety-critical medical devices can be traced back to more than one century ago, and similar to the history of IVEs, it begins with analogic devices and evolved to rather complex interactive computing systems. In 1895, Wilhelm Conrad Röntgen, accidently discovered the X-rays when manipulating a Crookes tube and within a year machines of the sort of the one in **Figure 56** (panel A) where being deployed in medical facilities. In terms of interface, these first generation medical devices relied on knobs and knife switches, and had little or no displaying of information (with the exception of gauges on early XX century diagnostic devices, see **Figure 56**, panel B).

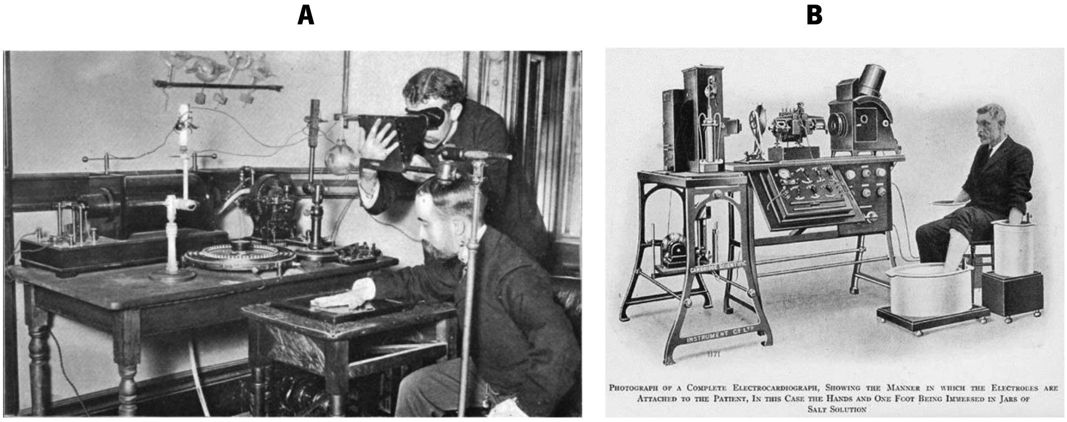

A
B
Photograph of a Complete Electrocardiograph, Showing the Manner in which the Electrodes are Attached to the Patient, In this Case the Hands and One Foot Being Immersed in Jars of Salt Solution

Figure 56. Early experiments with X-ray machines using Crookes tube apparatus (Panel A), image under Public domain - published in USA before 1923. An example of an early commercial electrocardiogram machine manufactured by Cambridge Instruments (Panel B), image under Public Domain.

With advances on the development of electronic components and a better understanding of the fundaments of bioelectricity, smaller programmable portable devices, such as the first wearable pacemaker (**Figure 57**, Panel A) became possible. In 1957, Medtronic started to commercialize wearable cardiac pacemakers that had controls to adjust the pulse to given amplitudes and durations, based on the particularities of each patient [168]. Today's wearable devices, such as portable infusion pumps, have numerous functions that allow the patients to define and schedule the values to be infused, monitor the progression of the treatment, and interpret complex logging data. These devices include wireless communication with hand-held peripherals and communication with software packages, allowing for storage of medical information and enhancing the process of therapy tracking (**Figure 57**, Panel B).

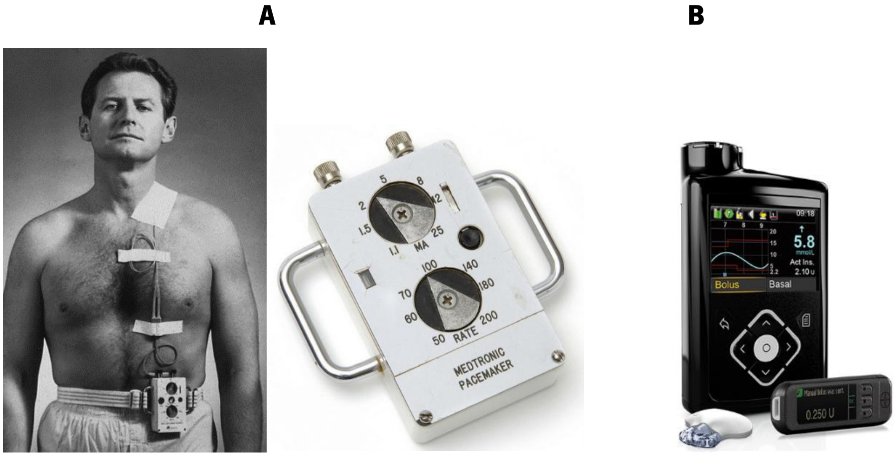


Figure 57. A Medtronic wearable external peacemaker from 1958 (Panel A), image from [168]. The Medtronic Minimed™ 640G an insulin wearable pump released in 2015, image from Medtronic, Inc.

Perhaps one of the best examples of the complexity of today's safety-critical medical systems is the *daVinci*® Surgical System from Intuitive Surgical, Inc. The daVinci is a teleoperation system where the surgeon controls four robotic arms, which allow for minimal invasion surgery through video-laparoscopy (**Figure 58**). These surgical systems are particularly challenging, in terms of HCI, because they combine the challenges of IVEs – surgeons rely on stereoscopic images of the intervened region [169] and AR systems were developed to assist during the medical procedures [170]– and the challenges of operating using typical controls of a safety-critical medical system.

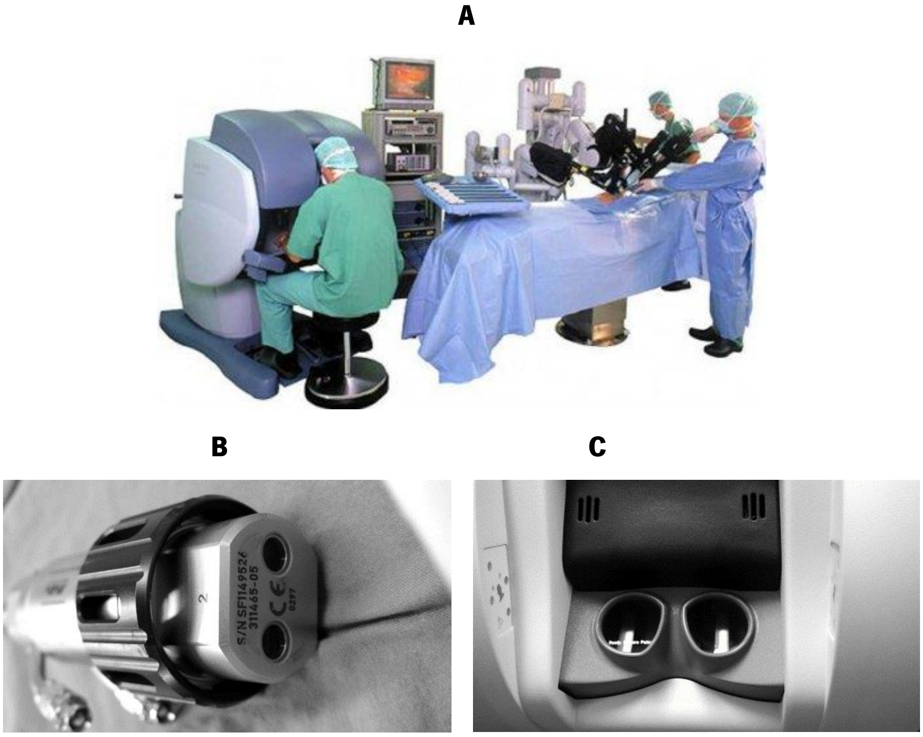


Figure 58. The daVinci® Surgical System from Intuitive Surgical (Panel A), image from Intuitive Surgical. A stereoscopic endoscope (Panel B) and the visualization apparatus from the daVinci® master console (Panel C), images from [Nam2012].

Modern medical devices constitute a good example of the complexity safety-critical medical devices can attain, and are thus a particularly relevant case study for the contribution analytical HCI methods can bring to the safety enhancement of these interactive computer systems. From a software development standpoint, safety-critical systems requires the application of carefully designed processes (often, standardized processes) during the phases of specification, architecture, development, and verification. There are two distinctive approaches during specification and design of these systems [171]:

1. Specify and design error-free systems, proving that faults are not possible – however, this approach only works for small systems, which are sufficiently compact for formal mathematical methods to be used in the specification and design, and with no or little intervention by a human operator. [172] presents one of the few examples of this approach, where a productive dialogue between the developers of a dialysis machine, with no experience or knowledge of formal methods, and computer scientists using formal analysis tools proved the efficiency and safety of certain software components;

2. To aim for the first approach, but to accept that malfunction, faulty modes, or use-errors might happen and to contemplate, during specification and design, error detection and recovery

capabilities. Due to the complexity of current safety-critical systems, this is the prevalent approach.

Thus, while there seems to be a high level of awareness among regulators and even developers for the need to apply standardized processes during development and evaluation of interactive safety-critical systems, as we will see, practice tells a different story. As Lyu famously exposed in his *Handbook of software reliability engineering*: "*The demand for complex hardware/software systems has increased more rapidly than the ability to design, implement, test, and maintain them.*" [173].

4.2. What is the Problem with Medical Devices Interfaces?

In the United States, when a medical device certified by the Food and Drugs Administration (FDA) is involved in the root cause of a medical accident, a description of the event, including causes that led to the medical incident, is included in the MAUDE[30] public database, and an investigation might be open to look into a possible recall of such device.

One of the most severe incidents registered, and one that sparked a lot of research in the causes of medical device use error, was the Therac-25 medical incident. The Therac-25 was a computer-controlled radiation therapy machine, commercialized in 1982 with the main purpose of delivering a highly concentrated dose of radiation to a part of the human body with minimum impact to nearby tissues. Taking into account the goals and context of use of such a device, one could guess that this is a technology to be used by skilled operators and that depends on highly safe software. Nevertheless, and contrary to what might be the patient perceptions, these safety-critical devices are often developed by small companies, at the time, with few established practices of safety-critical software development.

The Therac-25 was developed by a small team of engineers at a company called Atomic Energy of Canada Limited (AECL). While previous versions of the radiotherapy machine sold by AECL relied heavily on hardware control, this version of the Therac was the first generation to make extensive use of software control – developed by a single person in the company using PLP 11 assembly language and leaving the development process mostly undocumented. In a letter to FDA, the AECL admitted that the Therac-25 software evolved from the Therac-6 (an older machine from the same company) and that the current "program structure and software was using certain subroutines that were carried over to the Therac-25 around 1976" [174]. This migration of older software routines to a new machine and the removal of

[30] The Manufacturer and User Facility Device Experience (MAUDE) database from the FDA can be visited, here: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm

some hardware safety interlocks generated software “bugs” that led to two main operational faults [175]: (1) if the operator incorrectly selected one mode (the Therac-25 could deliver electron or X-ray beams) and then tried to change it quickly, this would set the beam to be delivered without the target (a structure placed in the path of the electron beans to direct its incidence) in place; (2) the electron beam could be wrongly activated during field-light mode, when no beam scanner was active or target was in place.

Both type of software errors could lead to overdosing the patient with radiation, nevertheless, it was a combination of special factors that gave rise to the first serious incident in 1985 – two years after the Therac-25 went into service. The East Texas Cancer Center (ETCC) was the place of two of the six major incidents with the Therac-25. Following the second event, Fritz Hager, a senior physicist at the ETCC carried out an extensive investigation that involved trying to reproduce the only feedback radiologist got from the device – a warning message (that did not appear in the manual) displaying “MALFUNCTION 54”. Upon contacting AECL, Fritz Hager was informed that “MALFUNCTION 54” meant that the Therac-25’s computer could not determine if an underdose or an overdose of radiation was delivered. Without much helpful information from AECL, the physicists at ETCC continued trying to replicate the malfunction and finaly succeeded. A description of the situation that triggered the malfunction is given in [175]:

> *Sometimes a user would select by mistake the X-ray mode (by pressing “X”) when the prescription would require an Electron mode. This was a common mistake because most treatments required X-rays and skilled operators were used to type “X”. To correct this, the radiotherapist at ETCC would use the “▯” key to enter edit mode and correct the “X” to “E” (Electron mode).*

However, what happened after the first mode selection was that the machine began to move a turntable that settled the target position for the X-ray beam, a process that would take about 8 seconds. While running the process of setting up the turntable, the Therac-25 would ignore any edit made during this time, thus a change to the type of treatment from a fast operator (capable of completing the setting up within 8 seconds) could trigger a given beam with the protective target in the wrong position. The reason AECL was unable to reproduce this error was that they were introducing the prescription data carefully and not in the typically fast way a skilled therapist would do it. Latter this malfunction was also found in older versions of the Therac (namely in the Therac-20), nevertheless there were no incidents reported because these versions had hardware safety interlocks – a fuse in the machine would blow when activating a radiation beam with the target out of place, thus stopping the treatment.

There are several important lessons to be learnt in terms of software and UI development from the Therac-25 incident, including:

1. Reuse of code from older versions, even when they have proven effective, is not a good practice since the changes in hardware and software of a new version might be incompatible with old software routines.

2. Hardware safety interlocks are important in these type of devices where redundancy eventually pays off.

3. Finally, and most important in the context of this thesis, is that when considerations about the users are not taken into account the occurrence of serious accidents due to use error is, sooner or later, quite probable.

In the case of the Therac-25, failure to consider the user's particularities happened in several instances. The error messages were impossible to be comprehensible to the operator and even the information provided by the company was insufficient to get awareness of the problem. However, and most importantly, the problem at the root-cause of the ETCC accident stems from a software defect that was only triggered by highly skilled behavior (i.e., the capacity to program a given prescription under 8 seconds), showing that this type of performance was not considered or replicated during tests and validation phases of this medical device.

According to [176], core software components responsible for human-machine interaction are of particular interest, since these modules are safety-critical in the sense that latent anomalies in their design can sooner or later lead to use error and potential harm. In fact, by analyzing the data on use-error with medical devices, one quickly realizes that the UI part of this type of safety-critical system often does not receive the careful consideration, during specification and design phases, which other components of the system do.

From a system-engineering standpoint, flaws in the Human-Machine Interaction (HMI) design often induce use errors[31]. In fact, investigations of incidents with medical devices usually reveal that HMI design flaws, rather than the lack of user training or inadvertent user behavior, constitute the main source of use errors [177], [178]. Thus, preventing use errors has been long acknowledged as a top priority in safety-

[31] *Use error* is an act or omission of an act that results in a different medical device response than intended by the manufacturer or expected by the user [202].

critical interfaces design [177], [179]. One of the first researchers to define the intricate relation between interface design and use error, was Jens Rasmussen, a Danish human factors researcher that pioneered studies on information processing in control rooms. In its seminal book, *Information Processing and Human-Computer Interaction*, Rasmussen stated that: "*Faults and errors cannot be objectively defined by considering the performance of humans or equipment in isolation*" [180]. This new posture, and incidents such as the Therac-25, opened the door for serious consideration on how the interface design can lead to use-errors and soon HCI and human factors researchers started to come out with DCMs capable of explaining the process of use-error and heuristics and design guidelines to try to mitigate it.

The proliferation of medical devices in harsh contexts of use, should require careful consideration of different types of users and different types of user behavior. The data-entry process of hospital infusion pumps is a highly studied process due to its proneness to use error with serious consequences. [181] describes a set of 53 use cases collected from the analysis of 16 medical devices from 10 different manufacturers, and from use-related adverse events reported in the FDA's MAUDE database. One of those use errors with a programmable infusion pump can be described as follow:

> *The key sequence 2 2 . 3 is erroneously registered as 2.3 or 0.3 without any warning or notification when the key sequence 2 2 is performed too quickly, resulting in user failure to enter the intended infusion parameter value.*

Once again, this use error happens when a highly skilled operator executes the task at a pace for which the UI design was not prepared for. The device fails to distinguish between consecutive input actions (e.g., double clicks) and single input actions made by the user, thus two consecutive fast clicks on the key 2 are not considered as so, due to an erroneous attribution of a key debounce. Interestingly, what could be a good safety interlock (the correct detection of a key debounce), can constitute a new hazard if it does not consider the data-entry times of highly skilled operators.

Portable programmable infusion pumps brought additional complexity to the problem of use error and data-entry (**Figure 59**). The fact that patients started to be allowed to define their own infusion parameters in ambulatory contexts, raised two new problems: the users for which the medical device ought to be designed are now much more diverse (e.g., regarding age, experience, and knowledge) and there is no way to control the context of use (e.g., no way to assure that the user is not interrupted amid the process of data-entry, and no way to assure that the user is in an environment where the auditory warnings are audible). **Figure 59** describes a known FDA recall of a portable infusion pump from 2014, due to an interface design error. Patients were accidently receiving an over-infusion of insulin because of unintended additional clicks on the DOWN arrow of this infusion pump when defining the value to be infused. When trying to reduce the amount of insulin, an additional click in the DOWN arrow at value 0.0 – that could simply be the result of involuntary motor variability of the user – would move the value to be infused from 0.0 (the minimum) to 10.0 units (the maximum), thus generating a dangerous over infusion. This use-error is a good example of how an inoffensive design choice for some data entry procedures – moving from the lowest to the highest value through wrap-around – can be potentially harmful in a safety-critical system.

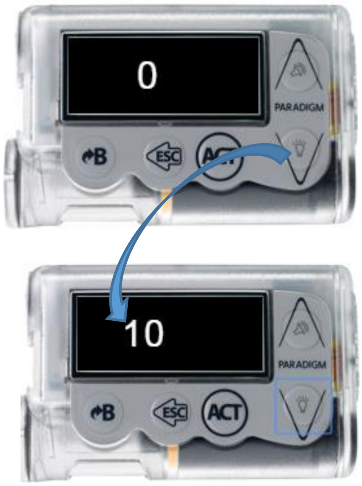


Figure 59. Minimized infusion pump from Medtronic, recalled in 2014 due to an interface design error. The device recall is registered here: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRes/res.cfm?ID=127259

From the examples presented in this section, it is clear that to design safer HMIs that prevent use errors, or facilitate the recovery from use errors when they occur, developers need to have a clear understanding of the relation between HMI design aspects, users' particularities and context of use. To better understand these relations, the role of DCMs can be crucial since this type of user models provide us with careful descriptions of the users' cognitive processes during interaction with technology.

4.3. Human Descriptive Cognitive Models

One DCM that had widespread influence, and it is still widely referenced today, is the Don Norman's *Seven Stage Theory of Action* proposed in 1988 [182]. Norman's seven stage theory of action defines the execution of a task as a set of iterations between physical system response and users' actions. According to this theory, any user action can be partitioned in seven stages (see **Figure 60**): 1) establishing the goal; 2) forming the intention; 3) specifying the action; 4) executing the action; 5) perceiving the system state; 6) interpreting the system state (i.e., making sense of the perceived information); 7) evaluating the system state (with respect to the goals and intentions).

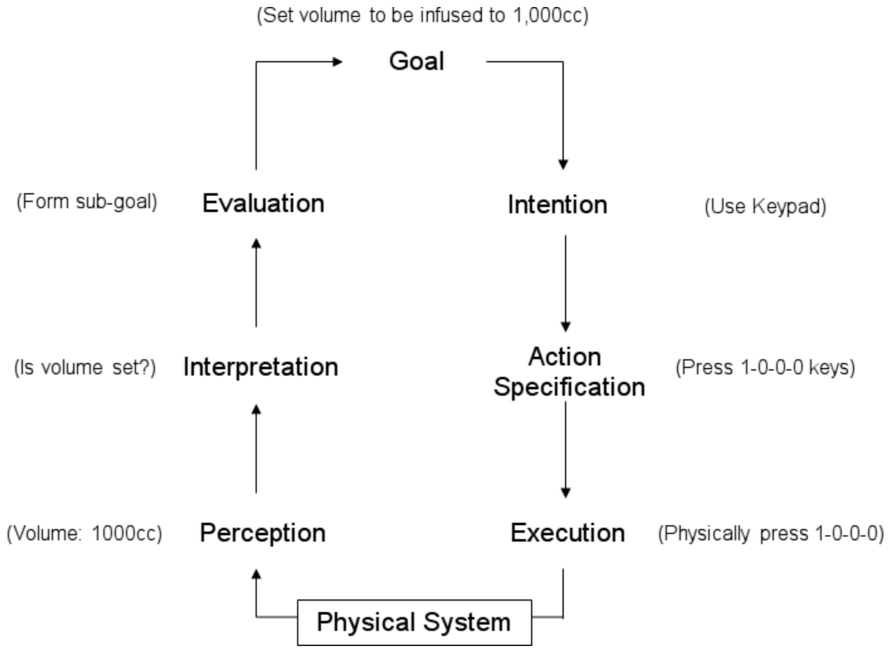


Figure 60. Norman's seven-stage theory of action. Examples on how to set the volume to be infused in an infusion pump is provided for each stage of the model (examples adapted from [183]).

The *Seven Stage Theory of Action* can be used to represent either a simple task with atomic actions or a complex task with nested sub-goals. It is of valuable use because it calls attention to the existence of different interaction stages with different cognitive demands and opens the door for considerations of great importance for developers of ICSs, such as: *what can go wrong in each interaction stage? what is the information relevant to mitigate use errors in each stage?* and *how might an error in one stage affect the subsequent stage?*

Nevertheless, and despite representing a single task or a complex task with nested cycles, the sequential nature of the action cycle described in Norman's model revealed itself unfit to capture the whole spectrum of human interactions with information systems. The experimental observation suggests that users tend to optimize the action cycle when they become skilled with the device (because of training, or repeated/frequent use) and that interaction resorting to the use of fewer stages is possible – for instance, upon perception and interpretation of a message a highly skilled operator can immediately execute the appropriate action. Jens Rasmussen recognized this limitation in Norman's model and suggested the inclusion of strategies that allow bypassing some of the cognitive stages. These strategies are called *cognitive shortcuts*, and are described in the DCM proposed by Rasmussen – the *Decision Ladder framework* [180]. This model was developed while Rasmussen was conducting research at the Risø National Laboratories in Denmark on interaction with nuclear power plants information systems. From analysis of decision making reports of operators, Rasmussen was able to extract the different information processing activities involved in different tasks, as well as some mental shortcuts adopted by the more experienced workers [184], [185]. Like Norman, he proposed a model with seven stages, but one major distinctive characteristic of the Decision-Ladder Framework is that it allowed to describe the interaction process as a passage from any stage to another, without having to sequentially transverse all the seven stages (see **Figure 61**). The existence of these cognitive shortcuts will vary according to the user and the heuristics the user has learnt through his experience with that particular device. Thus, differently from purely sequential models, the Decision-Ladder framework can account for different behaviors varying in complexity, which in turn depends on the context of use and on the experience of the user with the device. Although some researchers were already aware of the limitations of the sequential information processing paradigm [186], [187], it was only with Rasmussen's Decision-Ladder framework that the idea of cognitive shortcuts transformed the landscape in terms of applied cognitive models to human-machine interaction.

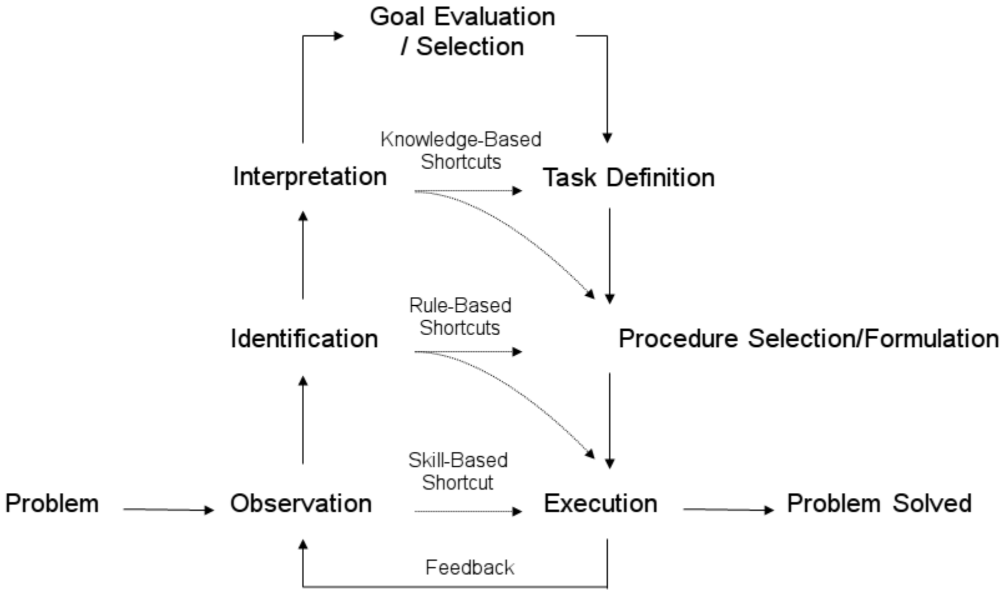


Figure 61. Decision-Ladder framework (adapted from [Rasmussen1986]).

In [188], the Decision-Ladder framework was further refined by identifying the cognitive shortcuts that are most important when analyzing human-machine interaction. These shortcuts are those that allow a rapid transition between three fundamentally different user's type of behavior and mental state:

(i) *Skill-Based Performance* – highly practiced, mainly physical actions in which there is no conscious monitoring. Typically used to complete familiar and routine tasks, skill-based responses are generally initiated by a specific event (e.g. alarm, visual cue, the need to start a routine task). Tasks performed with this type of behavior are so familiar that little or no feedback information is needed to accomplish the task.

(ii) *Rule-Based Performance* – in which the user applies a set of rules (learned through formal training, interaction with other experienced users) to a particular interaction. During rule-based performance the rules are being applied with minimal feedback from the situation, except perhaps for the detection of waypoints to indicate the correct progression of the rule-based procedure or sequence of actions [188]. The level of conscious control is intermediate between the skill-based performance and the knowledge-based performance.

(iii) *Knowledge-Based Performance* – is adopted when a completely new task or an abnormal situation is presented for which the user has no set of rules, and a novel plan of action needs to be formulated. In these situations, trained users tend to try to find an analogy between the unfamiliar situation and some known patterns of events for which rules are available in the rule-based level. Once a plan or strategy is developed using knowledge-based information processing, the user might revert to the previous levels to accomplish the task. Since it may involve a process of trial and error in order to optimize the solution, knowledge-based behavior involves a significant amount of feedback from the situation.

Because of its ability to account for different types of users in different contexts, the Decision-Ladder framework became the most widely used information-processing model for use error classification [189],with applications ranging from Lean manufacturing [190], to aviation and air traffic control [191], [192], to submarine navigation [193] and military land combat [194]. Today's practitioner's guides on process plants design and safe operation are still almost entirely based on Rasmussen's work (see [195]). Furthermore, one of the main frameworks for the study of complex socio-technical work systems, the

Cognitive Work Analysis (CWA) [196] uses decision ladders as one of its outputs. Recently two publications further analyzed the impact of Rasmussen's ideas on the disciplines of human factors, HMI, and HCI (see [197], [198]).

Both Norman's and Rasmussen's work have drawn attention to the fact that we should consider several cognitive steps that might be invisible to the external observer, as well as different types of performances when analyzing and describing the interaction process. James Reason, in 1990, built on this knowledge and focused his research on interaction failure, by developing a taxonomy capable of categorizing all the different instances of an interaction failure, thus creating one of the first use error taxonomies [199]. To design safer HMIs that prevent use errors and facilitate the recovery from use errors when they occur, developers need to have a clear understanding of the relation between HMI design aspects and use errors. A standard way to build this understanding is by using a use error taxonomy, such as the one proposed by Reason, which classifies use errors in accordance with systematic criteria [200]. Use error taxonomies are quite valuable for developers because they provide a reference to check what types of use errors typically occur during user-system interaction, as well as when these errors are likely to occur. In addition, a use error taxonomy can also help developers distinguish intricate differences between use errors stemming from different causes, and in turn devise effective measures to prevent and mitigate future occurrences (see **Section 5**). Reason's taxonomy, the *Generic Error-Modelling System* (GEMS), tries to explain how different types of performance (distinguished based on Jens Rasmussen's skill-rule-knowledge classification of human performance) relate with different types of use-error. Thus, GEMS guides the identification of three basic error types:

- *skill-based slips and lapses* - usually errors in highly repetitive, practiced tasks, due to motor variability (slips), or inattention or memory failure (lapses);

- *rule-based mistakes* – usually errors due to erroneous action plans as a result of picking an inappropriate rule or using a deficient rule, or caused by faulty rule-retrieving mechanisms coming from misconstrued view of the device state, an over-zealous pattern matching, or frequency gambling effects;

- *knowledge-based mistakes* - usually errors due to erroneous action plans as a result of incomplete/inaccurate understanding of system, confirmation bias, overconfidence, or cognitive strain.

According to Reason, Slips, Lapses and Mistakes may be distinguished based on certain dimensions, such as type of activity prone to their occurrence, attentional focus at the moment of use error, predictability of the use error occurrence, control mode of the operator at the moment of use error, ratio of error to opportunity of error, situational influences, ease of detection and relationship to change (see **Table 9**).

Table 9. Summary of the distinctions between skill-based, rule-based and knowledge-based errors (adapted from [199]).

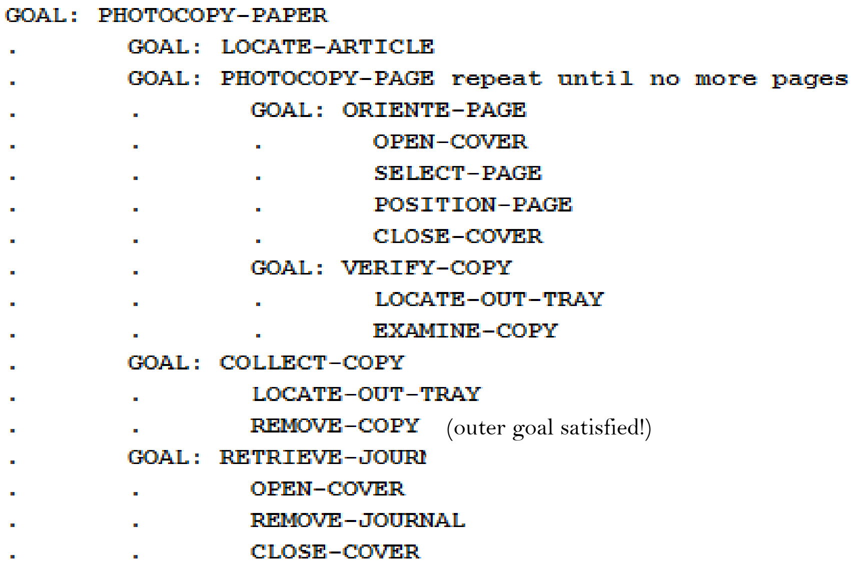
DIMENSION	SKILL-BASED Slips and Lapses	RULE-BASED Mistakes	KNOWLEDGE-BASED Mistakes
Type of Activity	Routine actions		Problem-solving activities
Focus of Attention	Out of the task	Directed at problem-related issues	
Control Mode	Mainly by automated processors (schemata or stored rules)		Limited, conscious processes
Predictability of Error Types	Largely predictable "strong-but-wrong" errors (actions or rules)		Variable
Ratio of Error to Opportunity of Error	Absolute numbers high but opportunity ratio low		Absolute numbers small but opportunity ratio high
Influence of Situational Factors	Low to moderate: Intrinsic factors (frequency of prior use) likely to exert the dominant influence		Extrinsic factors likely to dominate
Ease of Detection	Detection is usually fairly rapid and effective	Difficult, and often only achieved through external intervention	
Relationship to Change	Knowledge of change not accessed at the proper time	When and how anticipated change will occur unknown	Changes not prepared for or anticipated

Reason's use-error taxonomy became almost a standard for consideration of different types of use-error in generic tasks. Although it became widely applied in industries traditionally concerned with the effects of use-error (such as the aviation, medical, and automotive industry), the GEMS model stays true to its designation and often offers an overly generic set of classifications that fits all industries but stays short of capturing the complexity involved in each one.

The notion of the different types of behavior (skill, rule, and knowledge-based) and the consideration of different types of use-error (Slips, Lapses, and Mistakes), have become the cornerstone concepts of DCMs dedicated to describe the occurrence of use-error. Along the way some generic models which try to describe all the interaction cycle in all situations (faulty and normal interactions), have also been proposed. The GOMS model of Card, Moran, and Newell [70], is probably the most important examples of a general DCM. GOMS is an acronym standing for:

- **GOALS** – descriptions of what the user wants to achieve. Goals can be divided in sets of different sub-goals;
- **OPERATORS** – basic actions that the user must perform in order to interact with the system and ultimately achieve the goals;
- **METHODS** – different strategies that the user can adopt in order to achieve the goal;
- **SELECTION** – description about which methods will the user apply. This might depend on particularities of the user, on system state, or on details about the goals.

In a typical GOMS analysis, a high-level goal is decomposed into a sequence of unit tasks, all of which can be further decomposed down to the level of basic operators (see **Example 3**).



```
GOAL: PHOTOCOPY-PAPER
.       GOAL: LOCATE-ARTICLE
.       GOAL: PHOTOCOPY-PAGE repeat until no more pages
.       .       GOAL: ORIENTE-PAGE
.       .       .       OPEN-COVER
.       .       .       SELECT-PAGE
.       .       .       POSITION-PAGE
.       .       .       CLOSE-COVER
.       .       GOAL: VERIFY-COPY
.       .       .       LOCATE-OUT-TRAY
.       .       .       EXAMINE-COPY
.       GOAL: COLLECT-COPY
.       .       LOCATE-OUT-TRAY
.       .       REMOVE-COPY  (outer goal satisfied!)
.       GOAL: RETRIEVE-JOURN
.       .       OPEN-COVER
.       .       REMOVE-JOURNAL
.       .       CLOSE-COVER
```

Example 3. One possible GOMS description of the goal hierarchy for the task of photocopying an article from a journal (adapted from [10]).

This goal decomposition rational involves detailed understanding of the user's problem-solving strategies and of the application domain [10]. With their GOMS model, Card and colleagues, created a notation useful for describing the visible user's procedural knowledge when interacting with a system and this allows the analyst to easily grasp possible contexts for faulty interactions, dead-ends, or redundancy in the interaction flow. Looking at **Example 3** above, and as [10] pointed out, there is a post-completion error easily identifiable in the GOMS description. When the copy of the article is removed from the tray

[REMOVE-COPY], the outer (main) goal is already satisfied, which might lead to users moving away from the machine without fulfilling the last goal [RETRIEVE-JOURNAL]. In fact, this is a quite common use error – returning from a photocopier with the copy and forgetting the original in the scanner. As suggested by [10], one way of avoiding this could be a design that forced the goal [RETRIEVE-JOURNAL] to be satisfied before [COLLECT-COPY] becomes available, thus observing a good design principle – the obligation to satisfy all the mandatory sub-task before signalizing the completion of the main task. Today, software exists for automatic generation of GOMS-like models upon description of an interface and some rules of interaction (see *CogTool* [71] and *Cogulator* [72]).

DCMs are certainly capable of providing valuable information to developers, and ideally, they should be applied in earlier phases of design [12], as they are especially helpful as a preventive tool. Nevertheless, they can also be quite valuable as "forensic" tools to study what happened during a use-error event. The in-depth descriptions of the interaction processes DCMs provide are an ideal starting point to consider mitigation measures that can help to reduce or eliminate use-errors.

In **Chapter 4,** we presented the theoretical background that justifies the use of DCMs as efficient user-research tools during the development of new UIs. We have made the case that analytical approaches that take into account all the intricacies of the interaction process can become powerful tools in order to guide the development and analysis of UIs in safety-critical medical devices.

In the next chapter, we will describe the development of a new DCM (a new taxonomy for use-error in medical devices) and show how it can contribute for the development of safer and more usable safety-critical interfaces.

5. ANALYSIS OF HUMAN PERFORMANCE AND DEVELOPMENT OF DESCRIPTIVE COGNITIVE MODELS

"The human factor cannot be safely neglected in planning machinery"

Alfred Holt, 1878

In this Chapter, we will describe the development of a new DCM (a new taxonomy for use-error in medical devices) and show how it can contribute for the development of safer and more usable safety-critical interfaces. By doing this we will also demonstrate the role of analytical evaluation methodologies for the in-depth study of use-error in safety-critical UIs. We choose as a use case the safety-critical field of medical devices, where human error can have severe consequences and where the capacity to understand and describe the user interaction behavior is of crucial relevance. We will end this section describing the development of an IVEs for training with these safety-critical interfaces.

5.1. A Use Error Taxonomy for Improving Human-Machine Interface Design in Medical Devices

5.1.1. On the Importance of a Use-Error Taxonomy for Safety-critical Interfaces on Medical Systems

Medical Cyber-Physical Systems (CPS) typically incorporate a *Human-Machine Interface* (HMI) that serves as a centralized or distributed portal for users to monitor and control the system. Consider for example the Integrated Clinical Environment (ICE) [201], an interoperable infrastructure that coordinates multiple medical devices, medical apps, and other equipment to accomplish a shared clinical mission. ICE systems often provide a centralized HMI to allow the users to monitor and control the devices connected to the system. The HMI might also include safety interlocks and other safety systems (e.g., a centralized smart alarm system) to facilitate effective and safe user-system interaction.

It is thus critical to the safety of medical CPS to design safe HMIs that ensure expected user-system interactions and prevent potential *use errors*. Use error is an act of omission or commission performed by the user that causes a device to respond unexpectedly [202]. Preventing use errors has becoming a top priority in medical device design [177], [179]. From a system-engineering standpoint, use errors are often induced by flaws in the HMI design. In fact, investigations of incidents with medical devices usually reveal that HMI design flaws, rather than the lack of user training or inadvertent user behavior, constitute the main source of use errors [177], [178].

To design safer HMIs that prevent use errors and facilitate the recovery from use errors when they occur, developers need to have a clear understanding of the relation between HMI design aspects and use errors. A standard way to build this understanding is by using a *use error taxonomy* that classifies use errors in accordance with systematic criteria. Developers can use the taxonomy as a reference, to check what types of use errors typically occur during user-system interaction, as well as when the errors are likely to occur. In addition, a use error taxonomy can also help developers distinguish intricate differences between use errors stemming from different causes, and in turn devise effective measures to prevent and mitigate future use errors.

Numerous use error taxonomies have been proposed for medical systems with different degrees of specificity, scope, and coverage (see [200] for a survey). Recent years have seen a surge in frameworks to classify use errors with medical devices and better understand their causes in system design. In [203], Leveson's STAMP framework (*System Theoretic Accidents Models and Process*) is used as a basis to classify medical errors. STAMP is designed to support the analysis of causal factors not only at the level of unsafe actions committed by individual users, but also at management levels. While this broader view is certainly useful for healthcare providers to investigate system- and organizational-level causes of use errors, the error model used in the framework only coarsely classifies use errors into three categories: feedback, control action, and knowledge errors.

In [204], a *Human Factors Classification Framework* (HFCF) [205] from the avionics domain is adapted to the analysis of medical device-related incidents. The framework is built on Reason's GEMS [199] and similarly to the approach based on STAMP, HFCF considers use errors from a system-level perspective, and includes only five use error categories: decision errors, skill-based errors, perceptual errors, routine violations, and exceptional violations.

In [206] and [207], participatory design methods were used to create use error taxonomies for Computerized Physician Order Entry (CPOE) and tele-medicine systems. Participatory design builds on focus group discussions involving relevant stakeholders, including human factors specialists, cognitive specialists, social scientists and clinicians.

These taxonomies can be categorized in two main types: *model-based taxonomies*, which build on human cognitive models (an example taxonomy of this type is that proposed by Zhang et al. [183]); and *data-driven taxonomies*, which build on statistical data on use errors (an example taxonomy of this type is that presented in [208] for number entry errors). However, existing taxonomies have a variety of limitations that might result in incorrect, incomplete, or inaccurate classification of use errors. On the one hand, data-driven taxonomies often include ad-hoc error categories derived from statistics on medical incidents. They are usually limited to the current understanding and knowledge about use errors with a specific system. On the other hand, while model-based taxonomies in general promote a more systematic classification of use errors and their applicability typically extends across multiple device types [200], usually they build on Norman's *action theory model* [182], which oversimplifies the human decision-making as a sequential process with seven conceptual cognitive stages. Whilst Norman's action theory provides mental scaffolding for reasoning about the causal relation between HMI design aspects and use errors, the model disregards an aspect of human cognition that is important in the medical domain: *skilled behavior* due to well-practiced activities (e.g., learnt through training) or related to actions (predominantly motor actions) that can be performed with little conscious attention. This causes the taxonomies built on Noman's model to fall short when dealing with use errors committed by trained personnel, which is often the case with clinical operators of medical devices.

The Decision-Ladder framework describes human problem-solving as a seven-stage cognitive process, similar to Norman's action theory. As illustrated in **Figure 62**, these stages include: (i) *goal formation*, where one decides what needs to be done; (ii) *intention formation*, where one decides a strategy to achieve the selected goal; (iii) *actions specification*, where one identifies a concrete sequence of actions to implement the selected strategy; (iv) *actions execution*, where one performs the identified sequence of actions; (v) *perception*, where one monitors the effects of the actions; (vi) *interpretation*, where one develops an understanding of the perceived system state; and (vii) *evaluation*, where one decides whether the goal has been achieved. Failure to complete any cognitive stage is interpreted as a precursor to use error.

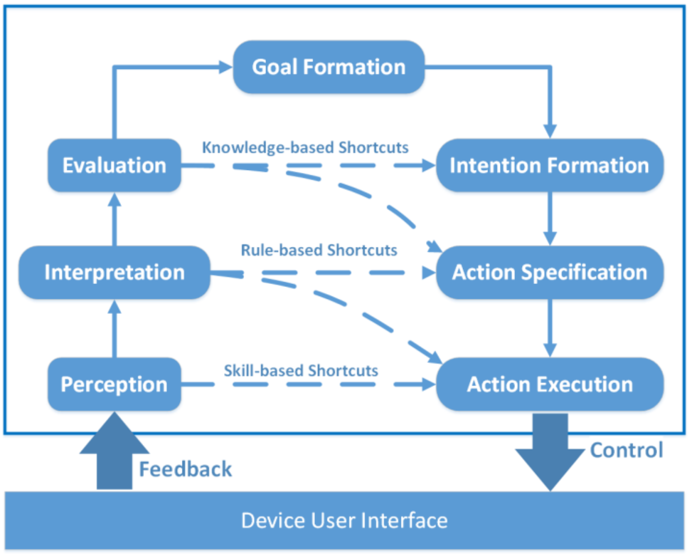


Figure 62. The Decision-Ladder Framework.

The advantage of the Decision-Ladder framework however lies in that it captures the case where one can traverse the cognitive stages in a non-sequential order, by taking certain *cognitive shortcuts* to skip cognitive stages. The cognitive shortcuts are useful for representing “rule of thumb” solutions adopted by experienced users when solving common problems [185], [188]. It should be noted that, when used appropriately, cognitive shortcuts can greatly improve interaction performance and also reduce use errors. The Decision-Ladder model includes three types of shortcuts:

- *Skill-based shortcuts*: heuristics adopted by skilled users when performing highly practiced actions during tasks (mainly motor tasks) – skilled users can complete these tasks with little or no feedback from the device.

- *Rule-based shortcuts*: heuristics adopted by trained users when performing procedural tasks they have learned through training or from previous experience – trained users typically rely on waypoints to monitor progress and status of procedural tasks.

- *Knowledge-based shortcuts*: heuristics adopted by experienced users when facing unfamiliar situations – experienced user tends to formulate an action plan by finding an analogy between the unfamiliar situation and some known patterns of events, and then execute the action plan (likely using skill-or rule-based shortcuts).

The use of cognitive shortcuts may vary across different users, depending on the heuristics they have learned and their past experience with a particular device.

5.1.2. A New Use-Error Taxonomy for Safety-critical Interfaces on Medical Devices

In this section, we present a use error taxonomy for medical devices that aims to address the limitations of existing taxonomies and explore the benefits of using a cognitive process model that is more sophisticated than Norman’s action theory model. In particular, we consider Rasmussen’s *decision-ladder framework* [180]. Thus, this work main contributions are: (i) the development of a use error taxonomy for medical devices to help developers reason about use error types and their relation with HMI design; (ii) an initial evaluation of the benefits of the developed taxonomy with respect to existing taxonomies.

We have followed the guidelines in [183] to develop our taxonomy:

- Step 1: Identify generic use error types by applying systematically a human error model to the cognitive stages of the selected cognitive model.

- Step 2: Elaborate an interpretation of the generic use error types with examples of the error types in the medical domain and typical HMI design flaws contributing to them. The elaboration is expected to better explain how the taxonomy is applied to medical devices and medical CPS.

Generic Use Error Types

To identify generic use error types, we build on Reason's *Generic Error Modeling System* (GEMS) [199], which is a de-facto standard approach. GEMS defines three generic error types:

- *Slips* - actions not carried out as intended;
- *Lapses* - missed actions due to temporary failure of concentration, memory, or judgement;
- *Mistakes* - errors due to erroneous action plans.

Applying GEMS to the Decision-Ladder framework is conducted by exploring the possibility of instantiating each error type at each stage and shortcut of the framework. This results in 16 use errors types (see **Figure 63**): 13 types of *information processing errors* due to failure to complete one or more cognitive stages; and 3 types of *cognitive shortcut errors* due to misuse of cognitive shortcuts during inappropriate situations. Note that not all GEMS error types can be applied to all stages. For example, slips or lapses are not applicable to the interpretation stage, as this stage relates to knowledge-based behaviors of the user.

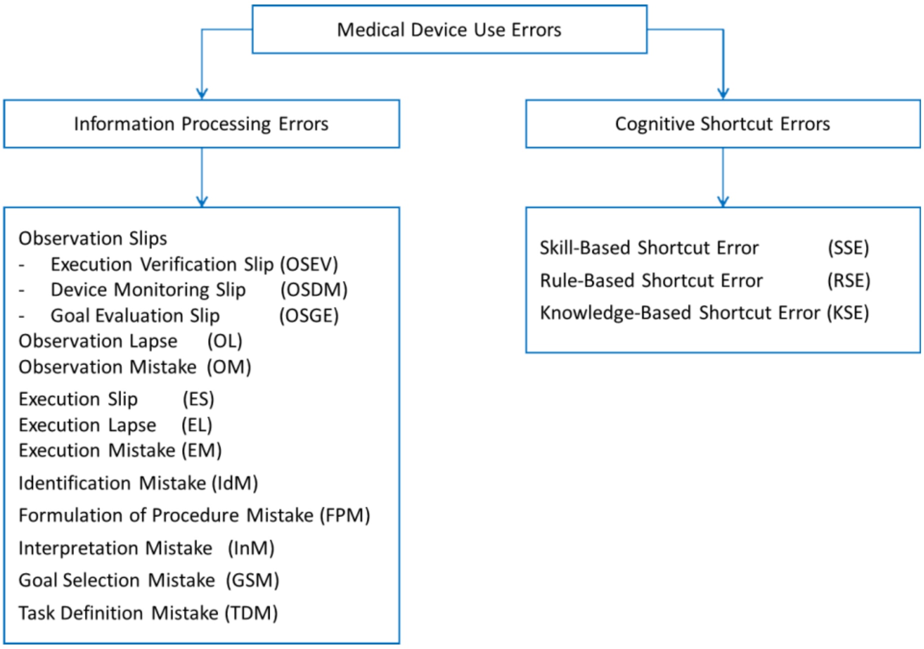

Medical Device Use Errors
Information Processing Errors
Cognitive Shortcut Errors
Observation Slips
- Execution Verification Slip (OSEV)
- Device Monitoring Slip (OSDM)
- Goal Evaluation Slip (OSGE)
Observation Lapse (OL)
Observation Mistake (OM)
Execution Slip (ES)
Execution Lapse (EL)
Execution Mistake (EM)
Identification Mistake (IdM)
Formulation of Procedure Mistake (FPM)
Interpretation Mistake (InM)
Goal Selection Mistake (GSM)
Task Definition Mistake (TDM)
Skill-Based Shortcut Error (SSE)
Rule-Based Shortcut Error (RSE)
Knowledge-Based Shortcut Error (KSE)

Figure 63. Use error types in our taxonomy.

We argue that the 16 identified types of use errors constitute a fairly complete classification of use errors because: (1) the Decision-Ladder framework covers both human decision-making activities and the user's expertise levels, and (2) the systematic application of GEMS to each stage and shortcut in the Decision-Ladder framework, wherever possible, enables us to cover human errors at any point of user-system interaction.

Medical Device Use Error Types

We tailored the generic description of the identified use error types to the medical domain by using information from medical incidents reported in the MAUDE database from the U.S. *Food and Drug Administration* (FDA) [209]. To this end, we have analyzed 50 incident reports involving use errors in 2014 and 2015.

The rest of this section presents the identified use error types, examples of medical incidents involving the error, and typical HMI design issues contributing to these error types. It is worth noticing that the HMI design issues (and examples) presented for the use error types are not meant to be exhaustive. Instead, they provide useful information for developers to better explore HMI design issues that likely contribute to medical device use errors. For the ease of reference, each error type is assigned with an error code.

Mistakes

Observation Mistake (code: OM): Failure to locate relevant feedback provided by the device, potentially resulting in observation or posterior use of information that is not relevant or suitable for the upcoming task. Typical HMI design flaws contributing to this type of error include:

- Information overload on the HMI;
- Lack of guidance, either on the HMI or in the training material, on how to access relevant information on the HMI.

Identification Mistake (**code: IdM**): Failure to identify which perceived device stimuli or information resource is important for the task being carried out. That is, the user is able to correctly perceive feedback from the device but fails to focus on relevant information. Typical HMI flaws contributing to this type of error include:

- Ambiguous presentation of information on the HMI (e.g., the device uses 'mg' as a shorthand for micrograms, whereas the user is trained to read 'mg' as milligrams);
- Lack of guidance, either on the HMI or in the training material, on how to identify relevant information on the HMI (e.g., a light is blinking to indicate some problem, but the user was unaware of the importance of that warning);
- Erroneous feedback on the HMI inducing confirmation bias (e.g., feedback on the HMI suggests data entry is completed when the device is still waiting for a confirmation action from the user).

Interpretation Mistake (**code: InM**): Failure to assign a correct meaning to correctly identified information, or to understand the implications of actions. Typical HMI design flaws contributing to this type of error include:

- Widgets on the HMI are labelled inconsistently across different device modes (e.g., the user is confused about how to cancel an action, as the soft key for the Cancel action is labelled as "Cancel" or "Back" in different device modes);
- The HMI provides incomplete information that causes the user to fail to understand the implications of their actions (e.g., the HMI display reports the infusion rate value without reporting the units during data entry).

Goal Selection Mistake (**code: GSM**): Failure to identify what needs to be achieved next. Typical HMI design flaws contributing to this type of error include:

- Information overload on the HMI (e.g., the HMI displays guidance instruction for the current action and the next required action at the same time);
- Incomplete feedback on the HMI (e.g., feedback on the HMI does not indicate that the device is in error condition);
- Incorrect documentation of what can be done with the device.

Task Definition Mistake (**code: TDM**): Failure to define a correct strategy to achieve the selected goal. Typical HMI design flaws contributing to this type of error include:

- The HMI does not provide appropriate functionalities necessary to accomplish required tasks;
- Lack of guidance on the HMI or lack of documentation on which device functionalities should be used to achieve a goal (e.g., a nurse fails to stop an infusion because it is unclear what sequence of actions should be used to stop the pump).

Formulation of Procedures Mistake (**code: FPM**): Failure to select the actions necessary for implementing the selected strategy. The difference between TDM and FPM errors is that the former are errors in deciding a strategy (e.g., adjust the dosage of the therapy to be delivered) whereas the latter are errors in deciding the actual sequence of actions (e.g., using the UP arrow key to adjust the dosage of the therapy). Typical HMI design flaws contributing to this type of error include:

- The HMI fails to respond to well-established user actions in a predictable way (e.g., the data entry system silently changes mode of operation in some boundary cases because, e.g., the keys for increasing/decreasing the infusion parameters become MR (memory recall) and MC (memory clear) functions in certain circumstances);
- Lack of feedback on the HMI or lack of documentation about how to address abnormal situation (e.g., error conditions). This flaw is particularly subtle, because it may result in mis-application of rules that are good in normal situations that are apparently similar to the present error situation but in reality are radically different.

Execution Mistake (**code: EM**): Failure to perform an action. Typical HMI design flaws contributing to this type of error include:

- Faulty HMI interlock that assigns the wrong priority to user actions that could be performed simultaneously (e.g., simultaneous presses of the 'stop' and 'retract' buttons is erroneously treated as 'retract');
- Lack of HMI interlocks that protect against dangerous actions during exploratory behavior (e.g., when the nurse is investigating ways to address an error condition);
- Lack of documentation or inappropriate HMI functions available to the user to perform required actions.

Slips

Observation Slips (**Code: OS**)**:** Involuntary error occurring due to failure to perceive device stimuli in different modalities (visual, auditory, haptic, etc.). Typical HMI flaws contributing to this error include:

- HMI design flaws in the way information is presented (e.g., the size, font, and spacing of letters do not allow clear reading of labels, such as in the case "VTBI19" and "VTBI9").
- Missing feedback on the HMI (e.g., a patient records screen does not provide sufficient information to allow unique identification of a patient, resulting in that clinicians accidentally derive information from the wrong patient or perform actions on the wrong patient record).

Observations slips are closely related to issues in device feedback, and can be further divided into three sub-types based on when an observation slip occurs in the interaction process.

Execution Verification Slips (**code: OSEV**). Observation slips occurring after the user executes an action. This type of slip might occur because the HMI fails to provide feedback in a timely manner when the user performs an action, or after the user has performed an action. This design issue can lead to cascading use errors because the user will tend to lose situational awareness about whether the device has actually registered the performed actions.

Device Monitoring Slips (**code: OSDM**). Observation slips occurring when the user checks the device state (e.g., failure to recognize that the device is running out of battery because the volume of audible alerts is too low to be noticeable).

Goal Evaluation Slips (**code: OSGE**). Observation slips occurring when the user checks whether a goal has been achieved (e.g., failure to understand if a therapy has been successfully started due to the lack of salient audio-visual feedback indicating such event).

Execution Slip (**code: ES**). Involuntary error occurring during the execution of an action. Typical HMI design flaws contributing to this type of error include faulty HMI design not taking into account one or more of the following situations: natural variability of motor actions performed by users; foreseeable accidental motor actions performed by users (e.g., the HMI erroneously wraps-around the rate value, and the user accidentally sets the rate to *max rate* rather than 0 because of an unintended additional click on the
DOWN button when the rate value is already 0); foreseeable data entry errors (e.g., typos, number inversions, pressing of multiple buttons at the same time).

Lapses

Observation Lapse (**code: OL**). These use errors are due to flaws in the user's encoding and memorization of device stimuli. They occur mainly due to inattention, memory loss, or interference. Typical HMI design flaws contributing to this type of errors include faulty HMI design erroneously delaying feedback without the user's awareness.

Execution Lapse (**code: EL**). These use errors are related to inattention, interruption, and/or intrusion into familiar patterns of activity. Typical HMI flaws contributing to this type of error include erroneous design of automated HMI functions that were meant to optimize interaction with the HMI, or handle exceptional use cases (e.g., the HMI silently discards data entry when the user pauses data entry for a period of time).

Shortcut Errors

Skill-based Shortcut Errors (code: SSE): Use errors during the execution of highly practiced tasks/actions (typically motor tasks). Typical HMI design flaws contributing to this type of error include:

- Faulty HMI interlocks that penalize expert behavior (e.g., the nurse quickly presses 2 2 ⊡ 3 and the device erroneously registers "0.3" instead of "22.3" because the first two key presses were performed too quickly and the HMI erroneously discards them as if they were a key debounce);

- Incorrect HMI layout for highly practiced motor actions (e.g., the HMI uses a numeric keypad with a phone layout in certain modes, and with a calculator layout in other modes).

Rule-based Shortcut Errors (**code: RSE**): Errors committed by the user during the execution of tasks/actions based on learned rules (heuristics). This type of error occurs when problem-solving strategies normally working in standard situations fail in the present context. Typical HMI flaws contributing to this type of error include:

- The HMI does not provide sufficient information to help the user reason about the implication of applying a learned rule in the present context (e.g., the HMI associates "slide left" to Confirm and "slide right" to Decline, which is opposite to typical designs, and does not provide visual cues indicating such non-standard association);
- Feedback on the HMI fails to call the user's attention to the inappropriateness of certain user actions in the current context (e.g., the data entry system always allows the user to enter an infusion rate with a fractional part, but only accepts it when the infusion rate is less than 100 mL/h. No feedback is provided indicating such constraint, or informing the user whether the entered rate can be accepted by the device).

Knowledge-based Shortcut Errors (code: KSE): Errors committed by the user during the execution of familiar tasks in unfamiliar situations. This type of use error involves the adoption of an erroneous problem-solving strategy based on stereotyped response to familiar systems states/modes. Typical HMI design flaws contributing to this type of error include:

- The HMI provides inconsistent functionalities in conceptually similar situations;
- Feedback on the HMI is not sufficient to discriminate whether the device is in normal operating conditions or in abnormal/error states.

Results: We applied our taxonomy to classify a set of medical device use errors with infusion pumps, reported by Masci in [181]. This collection of use errors with infusion pumps includes information from different sources such as: Incident reports collected through the MAUDE database; analysis of commercial infusion pumps and infusion pumps logs; data from workshops with manufacturers, users, and clinicians; and information from guidance documents (guidelines and standards). This set of use-errors allowed us to evaluate the quality of the taxonomy from two perspectives:

- *Applicability:* the completeness and ease of use of the taxonomy when applied to classify real-world use errors;
- *Accuracy:* the ability of the taxonomy in distinguishing use errors that are similar but their differences are worth highlighting (e.g., because they require different mitigation strategies).

A group of four experts (composed by a computer scientist expert in software engineering for medical devices; two FDA expert evaluators; and a cognitive scientist), were given the description of both taxonomies and of the set of use errors. At a first stage, they performed the evaluation independently and later they had two consensus meetings, where dissonant classifications were discussed and an agreement on each classification was reached.

Details of the data set of real-world use errors considered in our exercise can be found in [181], which includes 53 use errors collected from the analysis of 16 medical devices from 10 different manufacturers, and from use-related adverse events reported in the FDA's MAUDE database for infusion pumps, ventilators, patient monitors, and infant warmers in the decade 2000–2010. These use errors and their root causes have also been reviewed and confirmed by a team of device experts and healthcare practitioners.

Figure 64 illustrates the distribution of the considered use errors in our taxonomy (full details are available at https://goo.gl/eZAsRD). For each use error in the dataset, except for one, we were able to classify it to a single use error type. The only exception use error has the following description: *The user entered an incorrect key sequence "0 9" that was erroneously accepted and registered as "9" without any warning*. We classified this error as both Execution Slip (ES) and Skill-based Shortcut Error (SSE), because it can be interpreted as either an error committed while performing a motor action, or an error committed during a routine task. This exercise demonstrates that most use error type descriptions in our taxonomy offer sufficient guidelines in classifying real-world use errors, nevertheless it might be possible to refine some of these descriptions to better distinguish use error types that potentially overlap with each other.

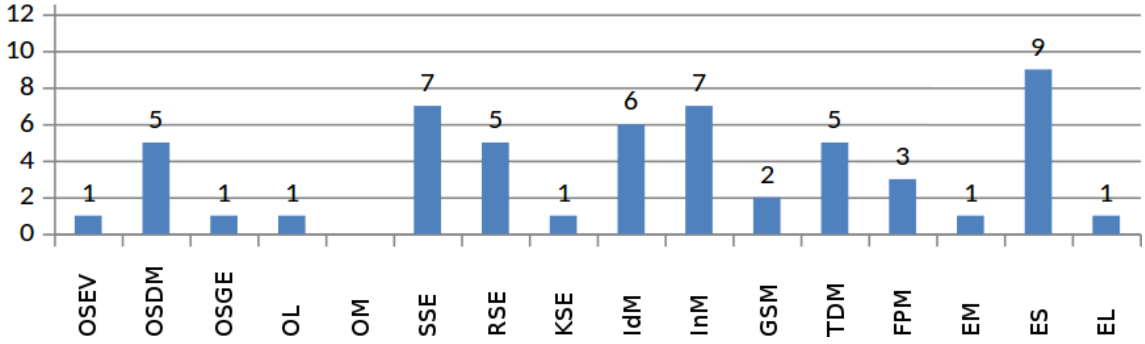


Figure 64. Classifying use errors with our Taxonomy.

Figure 65 shows the classification results using Zhang et al.'s taxonomy. We applied the taxonomy proposed by Zhang et al. [183], one of the most comprehensive taxonomies for medical use errors to date, to classify the same set of use errors, and compared its classification results with ours. This allows us to compare the classification accuracy of these two taxonomies. To ensure a correct and fair comparison, we strictly followed the definitions of error categorizes in [183] during the classification.

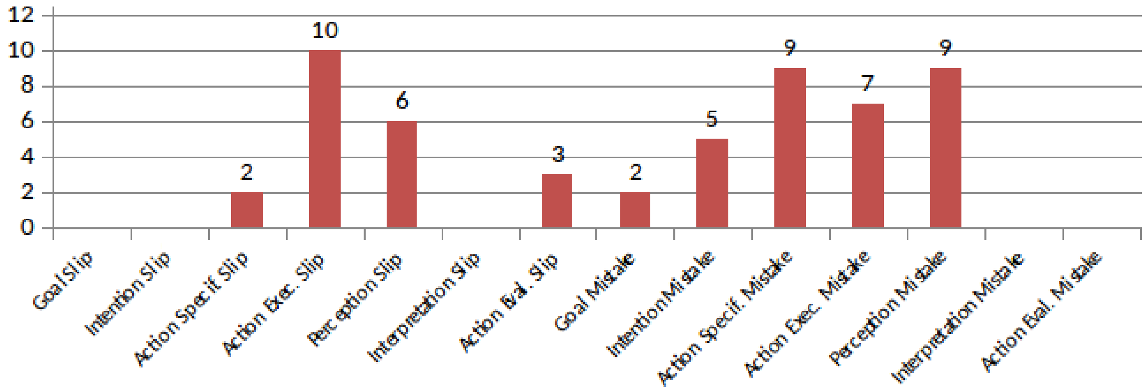


Figure 65. Use error classification with Zhang et al.'s taxonomy

Comparison between **Figure 64** and **Figure 65** indicates that use errors distribute more evenly in our taxonomy. It is also interesting to note that 5 out of 14 error categories in the taxonomy of Zhang et al. did not capture any use error, as compared to just 1 out of 16 with our taxonomy. This might indicate that our taxonomy has a finer level of granularity that enables more accurate classification of use errors.

During the comparison of these two taxonomies, we encountered two use errors that demonstrate the potential advantages of our taxonomy. As illustrated in **Figure 66**, two use errors were both classified as Action Specification Mistakes using the taxonomy of Zhang et al., while our taxonomy classified these errors as different use error types. In particular, our taxonomy classified one of these two errors as a Rule-based Shortcut Error (see the top of **Figure 66**), because the error was due to the fact that the heuristic adopted by the clinician was appropriate for similar clinical contexts but not for the current context. The other use error (shown at the bottom of **Figure 66**) was classified as an Interpretation Mistake by our taxonomy, because the HMI design flaw leads to a situation where the user is not provided with complete situational knowledge nor an understanding of the possible implications of the action.

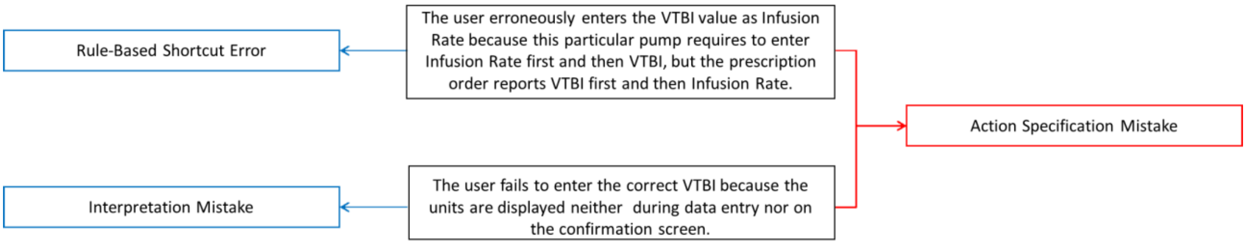
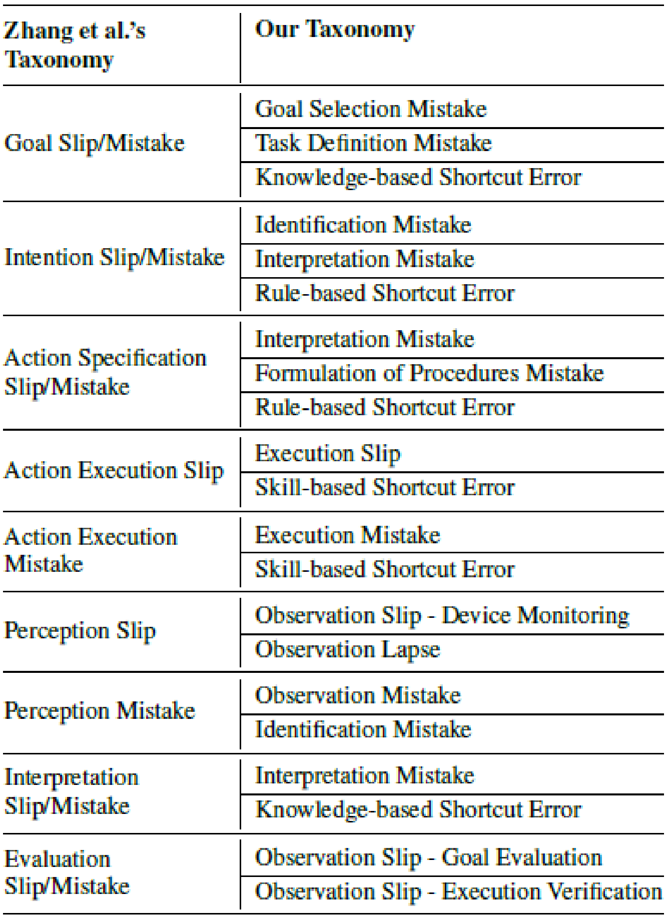

Rule-Based Shortcut Error
The user erroneously enters the VTBI value as Infusion Rate because this particular pump requires to enter Infusion Rate first and then VTBI, but the prescription order reports VTBI first and then Infusion Rate.
Action Specification Mistake
Interpretation Mistake
The user fails to enter the correct VTBI because the units are displayed neither during data entry nor on the confirmation screen.

Figure 66. Two examples of Use Errors and their classification using our taxonomy (left) and Zhang's taxonomy (right).

Classifying these two use errors to different categories can help developers better understand their root causes in HMI design and devise more appropriate mitigation measures. For example, the first use error can be mitigated by designing an HMI that accepts infusion parameters in the same order as that used in the prescription. The second can be mitigated by displaying the volume units next to its value during data entry.

Table 10 presents the alignment of our taxonomy with Zhang et al.'s. This alignment is particularly helpful to understand the difference between these two taxonomies and makes it easier to compare use errors classified using one taxonomy with use error reports classified using the other.

Table 10. Taxonomy alignment.



Zhang et al.'s Taxonomy	Our Taxonomy
Goal Slip/Mistake	Goal Selection Mistake
	Task Definition Mistake
	Knowledge-based Shortcut Error
Intention Slip/Mistake	Identification Mistake
	Interpretation Mistake
	Rule-based Shortcut Error
Action Specification Slip/Mistake	Interpretation Mistake
	Formulation of Procedures Mistake
	Rule-based Shortcut Error
Action Execution Slip	Execution Slip
	Skill-based Shortcut Error
Action Execution Mistake	Execution Mistake
	Skill-based Shortcut Error
Perception Slip	Observation Slip - Device Monitoring
	Observation Lapse
Perception Mistake	Observation Mistake
	Identification Mistake
Interpretation Slip/Mistake	Interpretation Mistake
	Knowledge-based Shortcut Error
Evaluation Slip/Mistake	Observation Slip - Goal Evaluation
	Observation Slip - Execution Verification

Use-Error and Design Solutions: Whilst being useful for classifying use errors with medical devices, any taxonomy should also allow device manufacturers to gain insights about how to improve their devices design. This can be done if there is a clear way to identify and relate each use error described in the taxonomy to a set of potential design solutions. The benefits of applying human factors principles to the design of a device was clearly demonstrated by Lin, Vicente, and Doyle [210]. They re-designed a Patient-Controlled Analgesia (PCA) pump that had been linked to several patient injuries and deaths, taking into account some well-established design principles during the development of their new prototype. Their results showed that the redesigned solution not only reduced mental workload and improved task completion times, but also reduced use errors by more than 50%.

Establishing a link between use errors and design principles might have a great impact in the mitigation of medical incidents. Based on Nielsen's usability heuristics and two usability standards currently in use ([177], [211]), we identified 12 design principles that could help mitigate use errors in different stages of the Decision-Ladder model (**Table 11**).

Table 11. A core set of interaction design principles from two usability engineering standards (ANSI/AAMI/IEC 62366:2009 and HE75:2009) and regulatory guidance documents.

No	Principle	Definition
1	**Visibility of status information and device events**	Relevant status information and device events should be available/perceptible on the user interface. This principle aims to keep the user informed about device events that require the user's attention, and operational modes that are necessary for correct decision-making when operating the device.
2	**Feedback to user actions and device events**	The effect of a user action or device event that changes the state of the system should be perceivable on the user interface. This principle aims to keep the user informed about what has been achieved.
3	**Salience of status information and device events**	Relevant status information and relevant device events should be prominent and easy to locate on the user interface. This principle aims to promote selective attention on important task-related information, providing the users with additional means that can help prevent oversight errors such as not noticing important status information or events reported on the user interface.
4	**Consistency of operating procedures**	The user should be able to operate the device in conceptually similar use context using the same operating procedures. This principle helps to prevent use errors due to controversial operating procedures.
5	**Consistency of device events**	Conceptually similar device events should be presented using the same user interface widgets. This principle helps to prevent use errors due to controversial interpretation of important device events.
6	**Consistency with user's knowledge and experience (Affordance)**	The user interface behavior should be consistent with the user's knowledge and experience. This principle limits the need for the user to learn new operating procedures and conventions, and helps to prevent use errors due to controversial action-effect mappings.
7	**Responsiveness to user actions and device events.**	Relevant information and events are reported in a timely manner during operation. This principle is important in clinical settings where the user can dedicate attention to the medical device only for brief periods of time, as they may become distracted by other tasks, or be interrupted during device use. In these situations, user interface software needs to be designed so that the user receives essential information about important events in a timely manner. However, software design should avoid imposing time limits for task completion, unless it is absolutely necessary, as different users and different clinical situations may require different response times.
8	**Reversibility of user actions**	Undo, redo, resume, restore should be supported by the user interface. This principle aims to allow users to easily stop, modify, and restart the automated processes controlled by software in the event they detect problems or abnormal situations.
9	**Forgiveness for use errors**	Adequate safety interlocks should be implemented to prevent foreseeable use errors, or help the user recover from errors. In particular, the software should be able to detect and report abnormal situations, and provide safeguards to ensure that the user cannot accidentally activate harmful controls.

10	**Predictability of automated user interface functions**	The behavior of automated user interface functions should be easy to anticipate. This principle helps the user to develop an accurate mental model of how automated user interface functions work, conveying information about whether automation is enabled, when it will be enabled and what user interface elements are controlled by automation.
11	**Simplicity of operational procedures**	Common operational procedures should require reasonably simple interactions. This principle aims to reduce the time and effort needed to complete common tasks.
12	**Traceability of user actions and device events**	Post-hoc reconstruction of user actions and device events should be supported. This principle aims to enable accurate reconstruction of human-machine interaction to support incident analysis, incident prevention, and forensic investigations.

In order to move up the Rasmussen's Decision-Ladder framework without errors the users have to correctly use the information they already possess. Thus, for instance, after correctly perceiving and identifying some device stimulus, the user needs to be able to assign a meaning to the device state/mode in order to fully interpret the available information. In the same sense, after defining a strategy to achieve the goal (task definition), the user should be able to define a sequence of actions that will allow him or her a successful execution (formulation of procedures). This interdependency between stages allows us to identify more easily what are the design principles that might help to inform the user at any given cognitive stage.

Table 12, lists exhaustively all the design principles that can give a contribution in successfully resolving that states of knowledge and help the user move up to the next information processing stage. Taking these design principles into account when designing the user interface of a medical device, does not guarantee the complete mitigation of use errors. Nevertheless, by applying these design principles developers are ensuring that the UI design provides the minimum information required for the user to resolve that information processing stage and move to other stages without eliciting a potential use error.

Table 12. Design principles suitable for mitigating use errors in the cognitive stages of the Decision-Ladder model.

DL Framework Stage	State of Knowledge	Potential Use Error	Design Principles
1. Observation	Perceive the outcome of an action.	OSEV	• Visibility of the outcome of user actions and device events. • Salience of feedback for user actions and device events. • Responsiveness to user actions and device events.
	Perceive device state/mode.	OSDM	• Visibility of device state/mode. • Salience of device state/mode.
	Perceive task completion	OSGE	• Visibility of task completion • Salience of task completion.
2. Identification	Distinguish the outcome of a different action	IdM	• Consistency of the outcome of user actions and device events. • Salience of the outcome of user actions and device events. • Simplicity of information indicating the outcome of user actions and device events. • Responsiveness to user actions and device events.
	Distinguish different device states/modes	IdM	• Consistency of device states/modes. • Salience of device states/modes. • Simplicity of device states/modes.
	Distinguish information indicating task completion from other type of information	IdM	• Consistency of information indicating task completion. • Salience of information indicating task completion. • Simplicity of information indicating task completion.
3. Interpretation	Assign a meaning to the outcome of an action	InM	• Affordance of the outcome for user actions and device events. • Simplicity of information indicating the outcome of user actions and device events. • Consistency of information indicating the outcome of user actions and device events. • Responsiveness to user actions and device events.
	Assign a meaning to the device state/mode	InM	• Affordance of device states/modes. • Simplicity of device states/modes. • Saliency of relevant information across device states/modes. • Consistency of information across device states/modes.

	Assign a meaning to information indicating task completion	InM	• Affordance of information indicating task completion. • Simplicity of information indicating task completion. • Consistency of information indicating task completion.
4. Goal Selection	Choose what needs to be done.	GSM	• Visibility of information indicating what needs to be done. • Salience of information indicating what needs to be done. • Consistency of information indicating what needs to be done. • Simplicity of information indicating what needs to be done.
5. Define Task	Define a strategy to achieve the goal	TDM	• Visibility of relevant widgets. • Salience of relevant widgets. • Affordance of relevant widgets.
6. Formulate Procedures	Define a sequence of actions	FPM	• Predictability of response to user actions. • Consistency of response to user actions. • Simplicity of response to user actions. • Reversibility of user actions.
7. Execution	Coordinate Manipulation	ES / EL / EM	• Consistency of response to user actions. • Simplicity of response to user actions. • Forgiveness to erroneous user actions.
8. Skill-Based Shortcut	Execution of highly practiced tasks/actions (typically, motor tasks)	SSE	• Forgiveness to erroneous user actions. • Salience of response to user actions. • Consistency of response to user actions. • Simplicity of response to user actions.
9. Rule-Based Shortcut	Execution of tasks/actions based on learned rules (heuristics)	RSE	• Consistency of response to user actions • Affordance of relevant widgets. • Simplicity of response to user actions. • Traceability of user actions.
10. Knowledge-Based Shortcut	Execution of familiar tasks in unfamiliar situations	KSE	• Predictability of response to user actions. • Affordance of relevant widgets. • Consistency of relevant widgets. • Simplicity of response to user actions. • Reversibility of user actions. • Forgiveness to erroneous user actions.

Discussion and Future Work: With this work we addressed the relevance of developing comprehensive and complete DCM. We have presented a use error taxonomy for medical devices with the aim of improving the understanding and awareness of all stakeholders on medical device use errors. The taxonomy was implemented in HTML and can be explored interactively here: https://github.com/Carlos-CCG/Thesis2019/tree/master/Taxonomy

The preliminary evaluation results of the taxonomy are promising, in that it allowed us to better distinguish use errors reported in medical device incidents as compared to existing taxonomies. This is critical for developers to understand the root causes of use errors and devise appropriate mitigation measures. Future work will concentrate on validating the taxonomy by applying it to larger datasets, and we will explore how to further refine the taxonomy to better distinguish similar use errors that warrant different mitigation strategies.

In the next section, we will go one step-further and show that while it is important to create DCMs that allow to classify use errors, it is also relevant to develop tools that allow practitioners to actively learn to avoid them.

5.2. Towards a Simulation-Based Medical Education Platform for PVSio-Web

5.2.1. On the role of SBME

As we could understand from **Section 5.1**, design problems and their contribution to use error have been acknowledged as a statistically relevant problem in medical devices incidents [212]. Walsh and Beatty, for example, estimated that 87% of medical incidents in patient monitoring contexts are due to human factors issues [213]; Brixey, Johnson, and Zhang estimated that use error has a greater impact in medical incidents than the impact caused by device failure [214]. Moreover, medical devices incidents suffer from severe under-reporting, with the total of reports ranging from as low as 1.2% to 7.7% of the total number of occurrences [212].

Medical incident reports are often the base to understand the root causes of use error with medical devices. Nevertheless, increasing the rate of use error reporting is challenging due to a propensity to assign culpability to the user (often the one reporting the error) instead of assign culpability to the device (i.e. faulty interface design) or the work environment (i.e. user interruptions or inappropriate workload management) [215]. Strategies to mitigate use errors range from systematics analysis of the device design, in order to detect and eliminate potential use problems, to practitioner training (e.g., in simulation environments).

The application of automated reasoning techniques to the model-based analysis of medical devices has been a topic of active research in recent years [216], [217], [218]. To better integrate the formality and rigor of the tool-supported analysis, with the iterative and exploratory nature of a typical human-centered design process, the need for simulating the models has been identified [219]. PVSio-web [220], for instance, is a tool that supports user interface prototypes generation from models of the user interface.

The prototypes can be used to validate the design with stakeholders, interpret analysis results and illustrate instances of use error, but also as a more economical alternative to actual device usage during training. PVSio-web generates prototypes supporting interaction with the device, enabling the user to explore the device behavior. The prototypes, however, are focused on specific devices, and do not provide a sense of the environment in which the device will be deployed. One way of dealing with this limitation is to create an IVE layer capable of framing the interaction with PVSio prototypes in an environment that is closer to their context-of-use.

Game-based learning is regarded as an emerging approach likely to have a large impact in training and education [221], and the use of simulators for skill training is widely regarded as beneficial [222], [223]. Gamification is one way to foster engagement during game-based training of residents' surgeons and in some instances, surgeons are required to practice on a weekly-basis in Simulation-Based Medical Education (SBME) platforms [216]. Nevertheless, current SBME platforms are costly and targeted at highly specialized groups of practitioners, such as surgeons [224]. SBME platforms targeted at healthcare practitioners such as nurses are still scarce, despite the fact that these professionals have to deal with complex medical devices often prone to use errors (infusion pumps, ventilators, and radiology devices) and, frequently, in hostile environments such as emergency rooms [225].

In this section, we present the requirements for the development of a new SBME platform based on PVSio-web and provide a first prototypical implementation of such SBME. The goal is that the proposed SBME contextualizes interaction with medical devices, such as ventilators and infusion pumps, in a game-based simulation of a hospital environment, thus improving the realism of the user interactions by simulating the context-of-use. To achieve this we added to the PVSio-web, the capabilities to simulate individual devices and the possibility to navigate in a virtual environment representative of the space where the devices will be deployed.

In this work, we will first set-up the requirements for developing an SBME, based on recently proposed design and development guidelines. Then, we will offer an overview of the proposed SBME and a description of its implementation. Finally, we will discuss the potentialities of SBMEs use for training of less specialized health care practitioners and the role that free, open-source platforms can have on fostering the use of such educational tools.

Requirements for the development of SBMEs

SBMEs platforms exist in a wide variety of technologies, in different degrees of complexity, and for different application purposes [226]. Regarding base-technology, SBMEs can range from paper-pen training toolkits such as the TeamSTEPPS [227], to virtual environments on desktop simulations (similar to computer games), high-fidelity immersive systems such as the training platform for the daVinci® surgical robot, or even real-environments on simulation centers resorting to real devices and surgical manikins. Regarding application, SBMEs might be designed to train professionals for a very specific procedure, such as a catheter insertion [228] or a knee replacement surgery [229]; to assess and score different professionals [230], to familiarize professionals with new medical devices before insertion on the ward; or even to foster team communication and other soft skills in different medical settings [231]. Despite all this variability, most SBMEs are applied as either a training or an educational tool, and this similarity in application scope led researchers on an effort to define requirements and key features of SBMEs.

Howe et al. [231] performed a review of work on SBMEs, and identified five key features for guiding the development of game-based simulation systems for medical training:

1. ***Automated Assessment*** – Use of scores or rating systems to evaluate users' performance. The assessment can be conducted through observational rating scales; self-assessments; or event-based approaches (scores systems implemented in the game);
2. ***Task Fidelity*** – Refers to the degree of similarity between tasks performed in the simulator and tasks performed on real equipment. Although the level of realism is something difficult to measure, the simulated environment should strive to mimic real-world interactions with objects, people, and teammates.
3. ***Interface Modality*** – Game-based simulators should allow for environment manipulation through user input. Some simulation events should be able to be controlled by the user and feedback of that interaction should be multimodal (sound, image, haptic feedback) and given timely.
4. ***Virtual Teammates*** – Simulations might encompass autonomous, animated agents that support face-to-face interactions. Virtual characters can serve as instructors guiding the narrative of the game, or playing roles such as team members, and thus promote the training of cooperation skills.
5. ***Customization and Adaptability*** – The training system platform must be flexible, customizable, and easy to use. Developers should be allowed to easily improve or expand previous implementations. This allows also for adaptations to each context of use by tailoring the simulation to represent the environment in a given hospital, or the use of specific equipment.

Other works are worth mentioning because they present similar and additional requirements for simulation environments. Issenberg et al., presents a list of features of high-fidelity medical simulations, ordered by prevalence in the literature [226]. It includes important additional features to Howe et al. features, such as:

6. ***Feedback*** – To foster learning, the systems should provide a way of informing users of their performance results. According to [226], educational feedback appears to slow decay of acquired knowledge while allowing learners to monitor their progress toward skill acquisition.
7. ***Multiple Learning Strategies*** – The SBME should be adaptable enough to be used in different educational contexts, such as instructor–centered education (lectures, workshops, educational videos, tutorials) or individual, independent learning without instructor.
8. ***Capture Clinical Variation*** – High-fidelity medical simulators that can capture or represent a wide variety of scenarios are more useful and allow for additional educational and training valences.
9. ***Individualized Learning*** – According [226], the SBME should convey the opportunity for learners to have reproducible, standardized educational experiences where they are active participants, not passive bystanders.

Taken together, these two lists offer a comprehensive set of nine requirements that should be considered when developing new SBMEs. We now analyse PVSio-web prototypes against these requirements:

1. Not being designed as a SBME platform, PVSio-web naturally does not support **automated assessment**.
2. Being based on models of the device, prototypes provide a high level of **task fidelity** at the device interaction level; however, they lack a more holistic view of the device in its environment.
3. Prototypes support user interaction and can support multiple **interface modalities**; e.g. voice output is supported.
4. Being focused on user-device interaction, PVSio-web prototypes currently lack support for **virtual teammates**.

5. The fact that prototypes are based on models makes them highly **customizable and adaptable**.
6. As with requirement 1, the prototypes were not designed with SBMEs in mind, so educational **feedback** is not directly provided; however, feedback about the interaction is provided as it would be in the real system.
7. Prototypes can be deployed in a number of ways so **multiple learning strategies** are supported and have indeed been experimented with.
8. Device prototypes are high fidelity and can **capture clinical variation**.
9. Prototypes are focused on user device interaction, so participants become active participants thus providing the basis for **individualized learning**.

In summary, two sets of features are needed to create SBMEs with PVSio-web: (1) Learning support features related to requirements 1 (Automated Assessment) and 6 (Feedback); (2) Features related to providing a more immersive experience of the clinical environment where the device (or devices) will be used in practice. This last set of features relates mainly to requirements 2 (Task Fidelity) and 4 (Virtual Teammates). Herein, we focus on the second set of features. We will do this by developing a virtual environment where PVSio-web prototypes will be deployed. This will add to the already realistic device prototypes a sense of the physical space where they are situated, without the need of using the actual physical space, thus improving task fidelity at a low cost. At the same time, it will also enable support for virtual teammates.

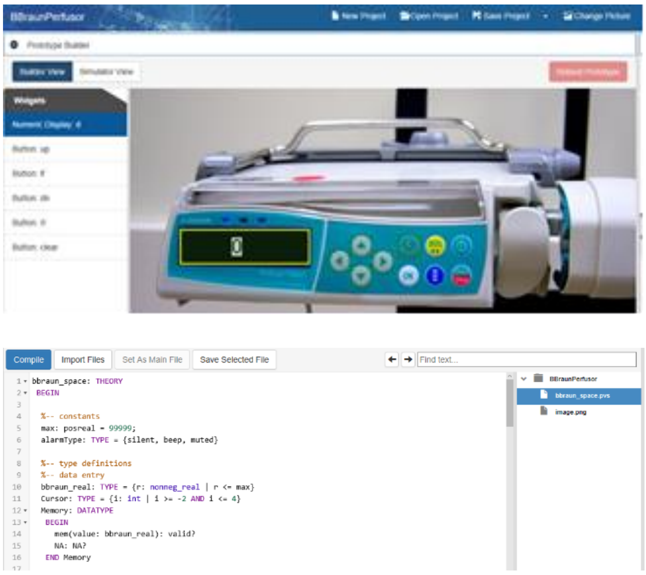
In the next sub-sections, we will offer an overview of the proposed SBME and a description of its implementation and. Finally, we will evaluate how well our proposed SBME complies with this set of nine requirements.

5.2.2. A proposal for a new SBME

Methodology

PVSIO-Web Description

The proposed SBME is composed of a game module that provides a structured, contextualized, and realistic use environment for the PVSio-web tool [220] – a graphical environment for facilitating the design and evaluation of interactive (human-computer) systems (se **Figure 67**). PVSio-web can be used to define, generate, and evaluate realistic interactive devices (existing devices or new concepts/prototypes) from formal methods. Since its release (in 2013), it has been mainly used to model and evaluate medical devices using formal methods, and to create training material for device developers and users in the medical domain – such as infusion pumps, and ventilators.



```
bbraun_space: THEORY
 BEGIN

  %-- constants
  max: posreal = 99999;
  alarmType: TYPE = {silent, beep, muted}

  %-- type definitions
  %-- data entry
  bbraun_real: TYPE = {r: nonneg_real | r <= max}
  Cursor: TYPE = {i: int | i >= -2 AND i <= 4}
  Memory: DATATYPE
   BEGIN
     mem(value: bbraun_real): valid?
     NA: NA?
   END Memory
```

Figure 67. Screenshots of the main tools provided by the PVSio-web environment. Top image – Prototype builder; Bottom image - model Editor and a model snippet generated from emuchart editors.

As stated in [220] state-of-the-art verification tools like PVS generally have minimal front-ends. This creates barriers for use by multidisciplinary teams and hinders engagement of non-experts. PVSio-web was proposed as a way of reducing these barriers and provide an interface to model medical devices. Our present work intends to take the tool a step further towards functioning as a SBME. We achieve this by developing a game that contextualizes and facilitates access to models available on the PVSio-web. Hence, the SBME we propose herein might be regarded as a *game mode* of the PVSio-web tool. At any given points in the game it calls upon interfaces prototyped in PVSio-web 2.3.

More precisely, the SBME is intended to provide a game-environment where navigation through different rooms of a hospital and interaction with different medical devices is possible. There is no pre-defined game narrative, thus free exploration of the game environment is possible and interaction with the different devices might follow any order. This environment is intended for external observers (evaluators or educators) to define their own script of interaction evaluation.

Simulation Software

To implement our SBME, we used Blender (release 2.78c) [47], an open-source 3D computer graphics software, written in C, C++, and Python, with an integrated game engine called Blender Game.

By choosing Blender as our simulation software, we are able to guarantee free, open-source distribution of a cross-platform SBME. Thus, other developers might be able to easily extend or adapt the game to each particular use scenario (i.e., different hospitals might require simulation of different medical devices). Moreover, by using free, open-source software we are keeping in line with the position adopted in the development of the PVSio-web tool.

Implemented Environment

The scenario for the game mode (see **Figure 68**) is a hospital ward, more specifically three adjoining rooms composed of a control room (Room 1), an fMRI (functional Magnetic Resonance Imaging) room (Room 2), and an operating room (Room 3). The player controls a character (a nurse) that can move between rooms and select the different devices available in the environment in order to interact with them. By default, the player starts at room 1, a control room for the fMRI machine. By opening the door with the fMRI machine off, the player can move to room 2, where it can view the interface of fMRI machines. By moving into room 3, the player enters an operating room where he or she may interact with several medical devices, such as infusion pumps and ventilators.

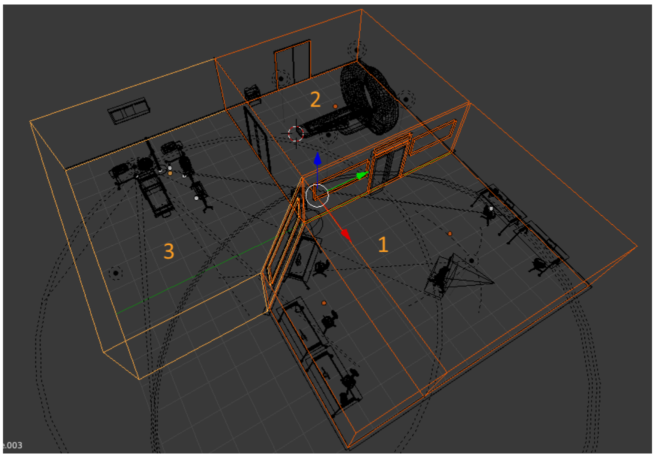


Figure 68. Game Scenario, consisting in three rooms.

Figure 69 presents the game flowchart and the adopted relation between the game mode and the PSVio-web tool in order to function as a SBME for training on medical device use.

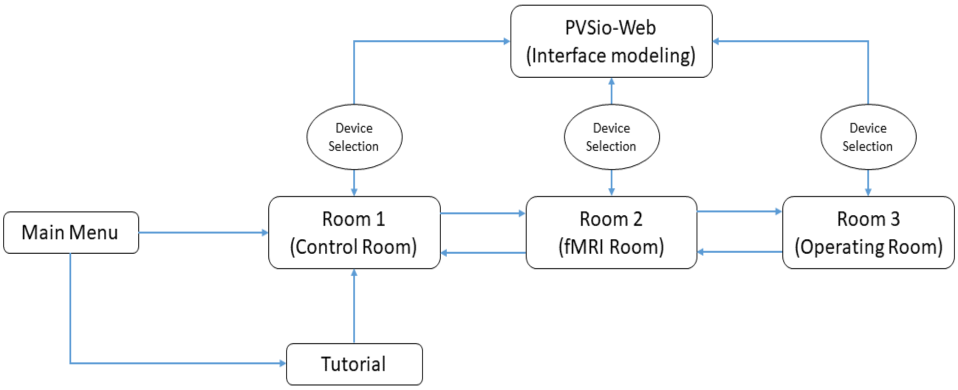
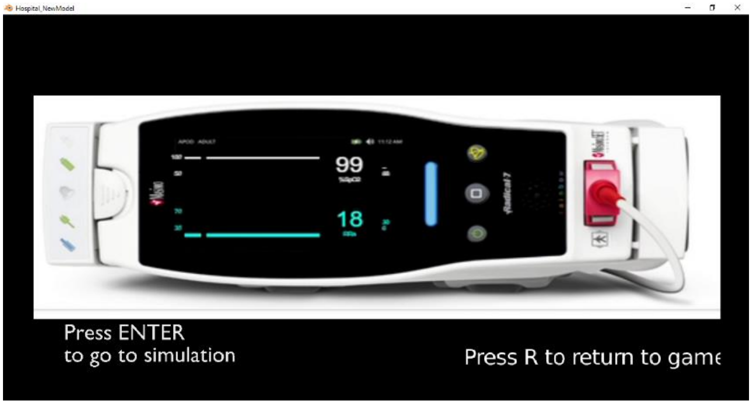


Figure 69. Game flowchart.

The transition between the game environment and the PVSio-web tool simulation is activated upon the selection of a specific device in a specific hospital room. Each time a player chooses to interact with a given medical device, the actuation on the main scenario stays on hold and a close-up of the medical device interface shows-up on the screen (**Figure 70**).



Press ENTER
to go to simulation

Press R to return to game

Figure 70. Close-up of the Massimo Radical 7® Pulse Co-Oximeter® a monitoring device that keeps track of patients' oxygen saturation.

To provide an environment familiar with a hospital ward, the rooms were modeled using high polygon count objects and high definition textures (**Figure 71**). The current total polygon count is 238 615. Next, we will discuss the main aspects of the environment. Namely, the available commands, the logging that is done, and concrete medical devices that were included.

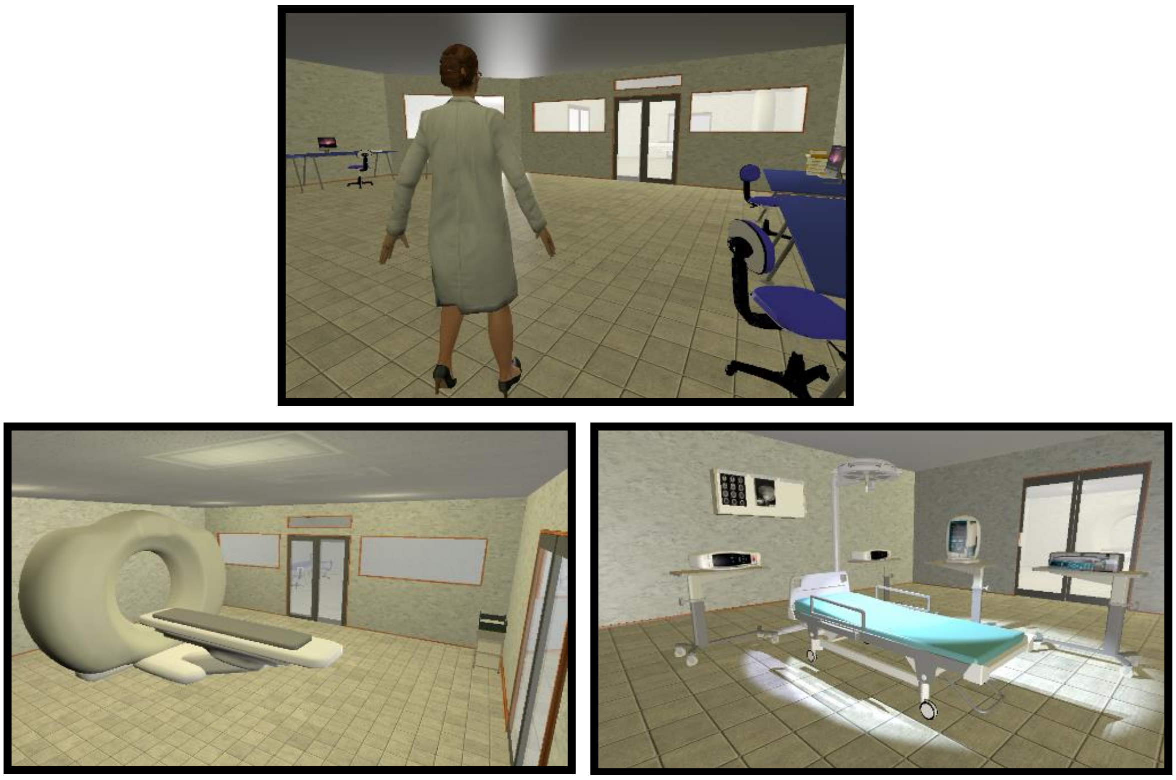
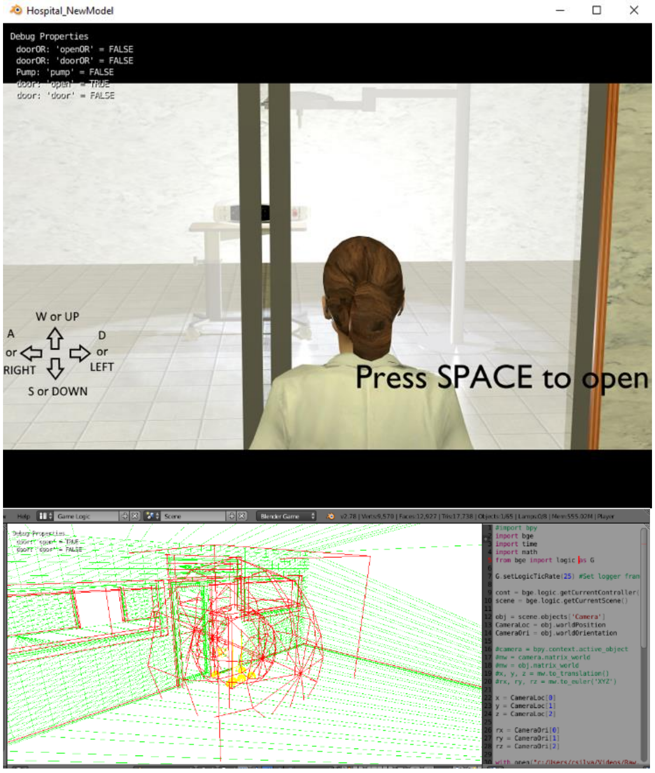


Figure 71. Screenshots from the game environment. Top-image is the control room; Left-botton image is the fMRI room; and Right-bottom image is the Operating Room.

Commands

The game features a tutorial explaining the commands available to interact with the environment. Namely, commands for character navigation control and commands for interaction with elements from the graphical environment. Character navigation is performed using W-A-S-D for up-left-down-right navigation respectively, and by using the mouse for camera pan and tilt control (resorting to the *look mode* on Blender's Game mouse actuator). Character control is implemented using the game logic mode of Blender Game.

Interaction with elements from the graphical environment was implemented resorting to Blender's *near sensors* and Boolean properties controlling animations (e.g. opening doors) or passage to new scenes (e.g. Medical devices close-ups). **Figure 72**, shows, through physics visualization, how the near sensor operates. A spherical area around the character is defined on the near sensor by setting a radius value to trigger the sensor (inward sphere) and a radius value to reset the sensor (outward sphere).

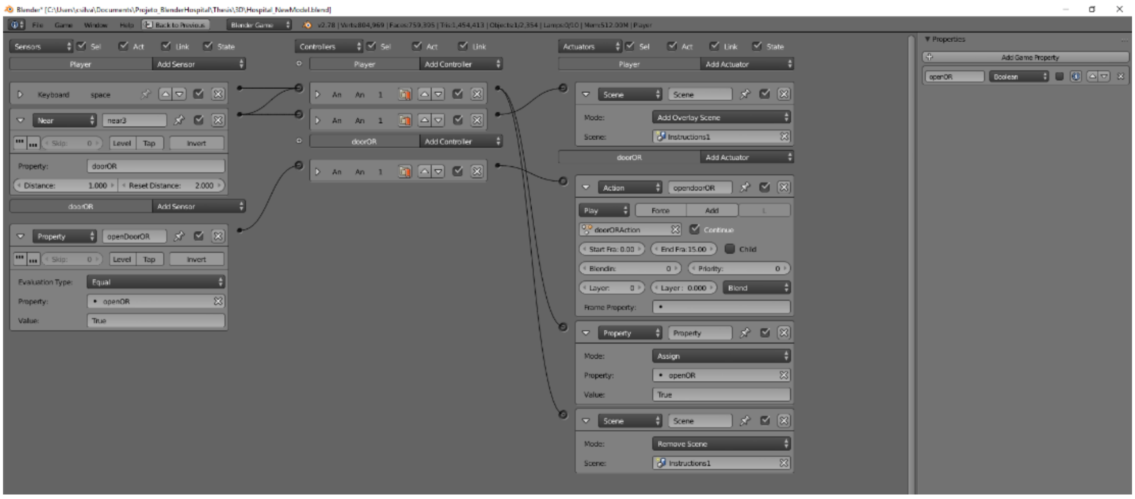


```
Debug Properties
  doorOR: 'openOR' = FALSE
  doorOR: 'doorOR' = FALSE
  Pump: 'pump' = FALSE
  door: 'open' = TRUE
  door: 'door' = FALSE
```

```
#import bpy
import bge
import time
import math
from bge import logic as G

G.setLogicTicRate(25) #Set logger fram

cont = bge.logic.getCurrentController(
scene = bge.logic.getCurrentScene()

obj = scene.objects['Camera']
CameraLoc = obj.worldPosition
CameraOri = obj.worldOrientation

#camera = bpy.context.active_object
#mw = camera.matrix_world
#mw = obj.matrix_world
#x, y, z = mw.to_translation()
#rx, ry, rz = mw.to_euler('XYZ')

x = CameraLoc[0]
y = CameraLoc[1]
z = CameraLoc[2]

rx = CameraOri[0]
ry = CameraOri[1]
rz = CameraOri[2]

with open("c:/Users/csilva/Videos/Raw
```

Figure 72. Top image, interface shown during the game in order to prompt the player to open the door. Down image, physical visualization of the near sensor in Blender Game.

Interactions on the SBME are restricted to opening of doors for room-to-room, character navigation, entering into close-up visualization of the different medical devices, and re-direction for the PVSio-web landing page of each device available in the SBME. Most of the interaction was programmed as a game logic in Blender Game. **Figure 73**, shows a simplified version of the game logic for the coordination between the *near sensor* and the *open door* interaction. A Boolean property is controlling the animation openDoorOR. A near sensor triggers the instructions in an overlay scene and sets the openOR property to true (if "SPACE" is pressed within the near sensor distance) allowing to run the animation of the door opening.



Blender* [C:\Users\csilva\Documents\Projeto_BlenderHospital\Thesis\3D\Hospital_NewModel.blend]

v2.78 | Verts:804,969 | Faces:759,395 | Tris:1,454,413 | Objects:1/2,354 | Lamps:0/10 | Mem:512.00M | Player

Keyboard space

Near near3

Property: doorOR

Distance: 1.000 Reset Distance: 2.000

Property openDoorOR

Evaluation Type: Equal

Property: openOR

Value: True

Scene Scene

Mode: Add Overlay Scene

Scene: Instructions1

Action opendoorOR

Play

doorORAction ☑ Continue

Start Fra: 0.00 End Fra: 15.00 Child

Blendin: 0 Priority: 0

Layer: 0 Layer : 0.000 Blend

Frame Property:

Property Property

Mode: Assign

Property: openOR

Value: True

Scene Scene

Mode: Remove Scene

Scene: Instructions1

openOR Boolean

Figure 73. A simplified Game Logic example for an opening door action.

Logs

For possible future analysis of the interaction and game performance, the environment is currently logging the character position throughout the entire duration of the game. This data might be interesting for analysis of the player's exploration patterns, or to quantify the amount of time the player spends in interaction with each device. The logger is implemented in Python through Blender's scripting platform. At each cycle of the defined logging frequency, the camera global location and orientation is printed-out in a .csv file. Dumping of the gameplay is also included as an option. By resorting to the *Rasterizer library* and using the *.makeScreenshot( )* function, we allowed gameplay dumping at a user specified frame-rate.

Medical Devices

The current version of the environment includes the medical devices described in **Figure 74**. Each device was modeled in Blender 2.78c (including the fMRI machine). These devices are shown in room 3 (see **Figure 71**). Once the player is near a device, an overlay scene instructs the player to press 'I' to view the device in a close-up mode (**Figure 74**). The close-up view shows a video of the device in running (for the pulse co-oximeter) or waiting mode (for the infusion pumps). This effect is achieved by UV-mapping an *.mp4* video as a texture of the object in the close-up scene. The game logic of that object uses an *Always Sensor* to run a python script that loads and refreshes the specified video into the object's texture, as follows:

```
IF “video” exists in the selected object
        “video” EQUALS to object’s “video”
        REFRESH “video”
ELSE
        READ object’s material and texture
        SET “video” variable as object’s video texture
        “video” source EQUALS to movie location
        IF “video” Loop parameter is false, make it true
        PLAY “video”
```

If the user, chooses to press “ENTER” when in close-up mode, the game is paused and a web-browser running that medical device simulation in PVSio-web is shown in full-screen. Inclusion of the PVSio-web simulation in the Blender game engine was not possible, since currently Blender does not support rendering of HTML inside a modal platform-native window as, for instance, Unity does (see Project Awesomium [232] for Unity). Keeping our SBME open-source was a requirement, thus we kept Blender as the development platform even though we had to deal with this limitation.

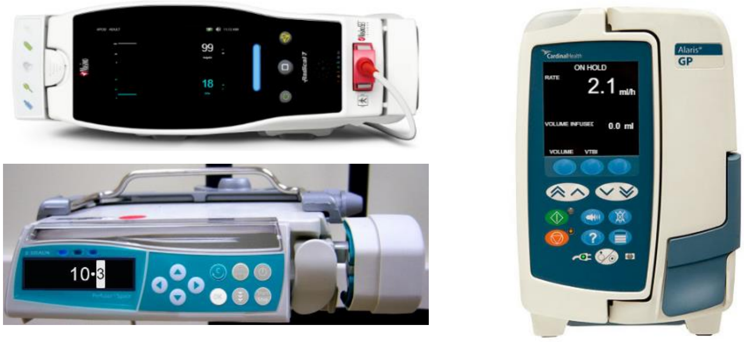


Figure 74. Medical devices currently modeled and available in our SBME. The top-left one is the Massimo Radical 7® Pulse Co-Oximeter® a monitoring device that keeps track of patients oxygen saturation, pulse rate, and perforation index; the bottom-left one is the BBraun® Infusionmat IV Pump, a programmable syringe based infusion pump; the right one is the Alaris® GP Volumetric Pump an infusion pump suited for drugs therapy, blood transfusions, and parenteral feeding

This project can be downloaded at:

https://github.com/Carlos-CCG/Thesis2019/tree/master/HospitalBlender

Evaluation: The current state of our SBME is already compliant with some of the requirements listed in Section II. Nevertheless, additional development is needed in order to fulfill all the nine requirements deemed relevant for game-based simulators in healthcare. We now consider the current state of development against the nine identified requirements:

1. ***Automated Assessment*** – Users' performance when interacting with the medical devices is not directly recorded by the SBME. Automated assessment could be possible if included in the models of the devices. This however is far from ideal. Currently, it is possible to evaluate the success of tasks completion through external observation or by resorting to interaction recording software for usability assessment, such as the OvoSolo® or the ActivePresenter7 software. In this sense, the SBME allows for *post-hoc* assessment of skilled interaction behavior. Future developments should consider the automation of the assessment process inside the game engine. As a side note, *post-hoc* assessment of the interaction with prototypes will be useful for usability assessment with end-users, during the development process of new medical devices.
2. ***Task Fidelity*** – By resorting to PVSio-web models our SBME allows interaction with medical devices simulations that mimic the exact behavior of real medical devices. Thus, learning transfer should be possible as the user might be prompt to perform tasks which are representative of the real-world scenarios. This is particularly important for training and assessment of performance on tasks that are prone to interaction errors (e.g. data entry in infusion pumps) or that highlight real faulty device behavior. Results of this interaction might serve, for instance, to raise healthcare professionals' awareness of the most probable use errors in each medical device. Additionally, the simulation now supports a sense of the environment where the devices are used and has the potential for implementing multi-user scenarios.
3. ***Interface Modality*** – Multiple modalities in the interaction with the simulated environment are possible. Nevertheless, the current implementation of the environment is still unimodal, lacking in auditory stimuli (unlike the individual device prototypes). Future developments should consider a multimodal simulation environment.
4. ***Virtual teammates*** – Although the environment supports developing multiple avatars, which might be controlled programmatically or by other users of the environment in a multi-user setting, currently there are no virtual characters serving as instructors or playing other relevant role in the scope of the developed environment. This feature might be relevant for future applications intended to approach training or education in more complex settings, such as the ones making use of cooperative work and medical cyber-physical systems.

5. ***Customization and Adaptability*** – By being developed in an open-source framework, our SBME allows customization and expansion by other developers. Moreover, recent developments in PVSio-Web facilitate the implementation of medical device simulations by non-experts in formal modelling [233].
6. ***Feedback*** – Due to the nature of its implementation, feedback on the interaction during a simulation trial is only possible if it is part of the formal description of the medical device simulation in PVSio-web. If the PVS model includes warnings, error messages, and state indicators, those will be visible during a simulation trial. Nevertheless, as the medical device simulation is developed outside Blender, this feedback behavior has to be implemented in PVSio-web. Additional educational feedback should be included in the game environment but it will be is dependent on the implementation of automated assessment (requirement 1).
7. ***Multiple Learning Strategies*** – Due to its nature as a platform for free interaction with the displayed medical devices, our SBME might be used in significantly different environments or applications. It can be used to develop educational videos illustrating the interaction problems that faulty implementations might create, to conduct controlled interaction assessment of healthcare professionals, or simply to provide a way for healthcare professionals to get familiar with the interface of a new medical device before implementation in the ward. Adding the simulated environment significantly increases the flexibility to implement different learning strategies.
8. ***Capture Clinical Variation*** – By being able to provide an interaction simulation with different medical devices, our SBME is able to account for some of the variability that healthcare professionals might encounter in real-world scenarios. Again, the simulated environment, significantly increases the expressive power of the tool.
9. ***Individualized Learning*** – By allowing reproducibility of scenarios of interaction and standardized educational experiences, and by demanding the users to be active participant during the simulation, our SBME might be applied in contexts of individual learning.

Table 13 presents a summary of the above discussion on the current state of the proposal as well as a comparison with the original PVSio-web prototypes, when analyzed against the requirements. The main conclusion is that support for requirements 2, 4, 7 and 8 has been improved. It should also be considered that the basis are established for improvements in the other requirements.

Table 13. Assessment: × - Not satisfied; ~ - Partially satisfied; ↗ - Improved; ✓ - Satisfied.

Req.	1	2	3	4	5	6	7	8	9
PVSio-web	×	~	~	×	✓	~	~	~	✓
Proposal	×	↗	~	~	✓	~	↗	↗	✓

Although some assessment is already feasible, the SBME should be expanded with a well-developed module for automated assessment of performance. By simulating contexts with standardized tasks, in which a particular action has to be performed in a particular device, the automated scoring of users' performance will be facilitated. As it is, the SBME allows only free exploration of user-interface interaction, requiring effort on external observers in order to keep track of users' performance. Nevertheless, a game logic with specific tasks and classifiable outcomes can be developed and interaction performance can be better assessed in future developments.

One of the main current advantages of our SBME is the possibility to customize and adapt the simulation wich, might be further enhanced by developing structured ways to conduct such customization. One way of doing so, is to control the process of adding new medical devices. Currently, this process is time consuming. The developer has to: (1) implement a formal description and a model for simulation in PVSio-web and, (2) create a 3D model in Blender with the game logic, which allows the player to see the device in close-up mode and progress to the landing page of the device simulation in PVSio-web. Currently, part (2) is already implemented as we offer an assets library in Blender, with a preview module for each asset, which allows the SBME scenario creator to easily select previously developed models of new medical devices.

Discussion: In this section, we present the basis for a free, open source, SBME for training of user-interface interaction in medical devices. Typical SBMEs are expensive, difficult to adapt to new use cases, and targeted at very specific sets of medical specialties [222]. The current state of our SBME provides the basis for developing training programs targeted at specific use cases (i.e. a hospital and their specific device models), and for less-specialized healthcare professionals, such as nurses. This particular group of healthcare professionals has to interact with an increasingly different set of medical devices, sometimes having to work in the same hospital with different models or brands of equipment targeted for the same function. All of this variability in medical device use hardens the process of learning and makes training even a more important part of the job. In a 2015 survey [225], 76% of respondents (from a pool of 113 resident nurses) stated the need of continued practical training. An expandable and easily adaptable SBME might configure a possible solution for this acknowledged necessity.

5.3. Conclusion

In **Chapter 5**, we first presented the advantages of developing new DCMs and of their application to a specific group of HMIs (the medical devices UIs). Moreover, we also discussed the role of IVEs as potential training platforms to mitigate the incidence and impact of use error in safety-critical interfaces and we proposed a first version of a free, open source, SBME for training of user-interface interaction in medical devices.

In the final Part of this thesis, we will summarize the work presented in this thesis and discuss its impact in the development present and future ICSs.

Part IV

General Conclusions and the Future of ICSs

6. GENERAL CONCLUSION AND THE FUTURE OF ICSs

> *"Powerful new aspects of the future are all the more persuasive when situated in existing, familiar experiences that we value and want to preserve"*
>
> Nick Montfort, *The Future*

When Douglas Engelbart, the famous HCI pioneer (see **Section 1.1**), died in 2013 the headlines announced that the computer mouse inventor had passed away [4]. Although there is no small achievement in being proclaimed the computer mouse inventor, one can guess that throughout his life, Engelbart could rightly feel the injustice of the lesser part all his other inventions have played in the public's mind. Although in his "mother of all demos", he presented the first materializations of other interesting tools such as video-conferencing, collaborative text-editing, and hyperlinks, it was the computer mouse that first appeared in commercialized systems and rapidly spread as one of the main user interfaces in ICS[32].

The explanation for the computer mouse success has to do not only with the fact that it is a clever way of "...*connecting the flat surface of a desk to a computer's planar grid*" (p. 90) [4]. Indeed, part of its success has to due to the fact that, when it was presented to the public, the computer mouse was already part of a much larger system which had a clear purpose of helping its users accelerating or increasing performance in work related tasks. Its role was easy to understand and its benefits, both for the overall system as for the outcome of using said system, were clear and even measurable.

The same principle should apply to any ICS and particularly to the two presented in this thesis: IVEs and safety-critical devices. To a great extent, the success of any ICS is dependent on how well the new system works as an effective tool during entertainment or work related activities of its users. In this regard, the primary goal of this thesis was to show how the two traditional approaches to user assessment in HCI – empirical and analytical approaches in the form of predictive and descriptive models of the user, respectively – could still be quite relevant for tackling the challenges of new ICS that might affect their adoption and widespread use.

One first question we intended to help answer was, how will IVEs ever attain the usability and user-acceptance levels of the traditional ICS paradigms? It was our argument, throughout this thesis, that to

[32] Even the computer mouse had a troubled journey from development to spread adoption. In the late 70's, the *Augmentation Research Center* run out of funding and many of its researchers and engineers migrated to Xerox PARC alongside with some of Engelbart's prototypes. At Xerox the graphical user interface and the computer mouse were pitched to the administrative board, without success. In 1979, the Apple team visited Xerox PARC and got hold to their GUI principles, which later were extensively applied to their first two commercial system, the Apple Lisa in 1982 and the Apple Macintosh in 1984. Thus, it passed 14 years since Engelbart's "mother of all demos" and the inclusion of the computer mouse in a commercial ICS.

foster the reality of IVEs as credible UIs we should first grasp the principles of how to measure perception and interaction with these systems and how to feedback these empirical observations to their development process, throughout new PCMs tailored to the IVEs context. Thus, in **Chapter 3,** we presented IVEs that were purposely developed as highly controlled experimental scenarios allowing for the assessment of users' audiovisual perception. The outputs of these experiments are PCMs quantitatively describing different perceptual phenomena that are crucial for a correct use of audiovisual IVEs. From the experiments described in **Section 3.1** to **Section 3.4**, we were able to extract the following guidelines:

Table 14. Summary of the IVEs design guidelines outputted from the experiments in **Chapter 3.**

User Perception of...	PCMs and Guidelines for IVEs Developers
Audio-visual Synchrony	• End-to-end delays of visual and auditory output should be measured and controlled in a way that a visual and an auditory stimulus modeled to be perceived synchronously, should not lag more than 62.5ms (half of the average WTIs value); • On an audiovisual synchronic scene, the vision-first bias should be taken into consideration and sound should never precede image; • The natural occurring sound lag should be taken into account and modeled when developing the IVEs. The rule for a buffer responsible for defining the sound delay should be: $\boldsymbol{y = y_0 - k + (3.4x)}$, *where* y_0 *is the measured difference between the audio and the visual end-to-end delay of a specific IVEs, k is a constant delay inserted to compensate for this difference, and x is the distance-to-user of the simulated audiovisual event.*
Audio-visual Unity Assumption	• Developers should avoid audio-visual spatial mismatches that render the audio at a position closer than 9.5° to 12° azimuth from a conflicting visual stimulus; • When possible, co-localized visual stimuli should be used in order to improve spatial discriminability of auditory stimuli;

Visual Distance Perception	• The FMDT is an appropriated task to measure the validity of IVEs requiring the performance of distance judgments by the users. • Real-world distance judgement follows a linear function with a slope higher than one, meaning that the natural occurring distance judgement error increases with the distance of the visual stimulus. In our study the PCM for real-world distance judgments is: $\boldsymbol{y = 0.86 + (1.1x)}$, *where x is the physical distance of the visual stimulus in meters.* • Photorealism minimizes the discrepancy between performance in simulated environment and real world, probably by reducing IVEs user's detection of non-identities.
Auditory Location	• Choosing display or audio output devices should be guided by evaluation of users' perception in controlled conditions; • Open-cascade headphones, which allow individualized pinnae and ear-canal modulation over the non-individualized HRTFs, are preferable over listening devices that have in-ear sound output. • When using non-individualized spatial cues, a short training session is advisable and can result in considerable improvements in location tasks. However, for the training improvements to be transferrable from training sessions to generic use situations, one should maintain the displaying or audio output device.

Throughout **Chapter 3**, we adopted the empirical approach that characterized HCI in the 80's, and applied it to audiovisual perception phenomena that we deemed crucial for a correct – and coherent with the real world – perception of a virtual stimulus. Our experiments were mostly concerned with the temporal and spatial features of virtual stimuli, because we deemed the correct simulation of those as the basis for any IVE intended to achieve satisfactory levels of fidelity. The implementation of these experiments, that we made available, are easily replicated in different virtual reality systems (both projection-based and head-mounted) and thus, taken together, they effectively serve the role of a user **perceptual assessment toolkit.**

Although the recent success of IVEs foretell additional applications in the near future, this type of environments is not yet prone to be used as a traditional work interface – i.e., interfaces where users can perform skill or rule-based tasks [180] involving issuing of specific commands such as data entry, activation of specific functions, and monitoring of relevant information. DCMs are particularly suitable for the study of interfaces used in routinely tasks (such as work and production tasks) and to identify instances of use error, thus we applied this analytical approach to our user-case of medical devices. In **Chapter 5**, we showed how one could derive a use error taxonomy (**Figure 75**) from a comprehensive descriptive model of human-device interaction, and we demonstrate that our use-error taxonomy appears to be more complete and granular than the current state-of-the-art ones. With this chapter, we intended to both call attention for the role of analytical approaches in interface design, while showing to developers how to build understanding of the root causes of use errors and devise appropriate mitigation measures. Furthermore, a link between use errors and design principles (based in Nielsen's usability heuristics [12] and two usability standards currently in use, [177] and [211]) that might help to mitigate them, was articulated in **Section 5.1.2.**

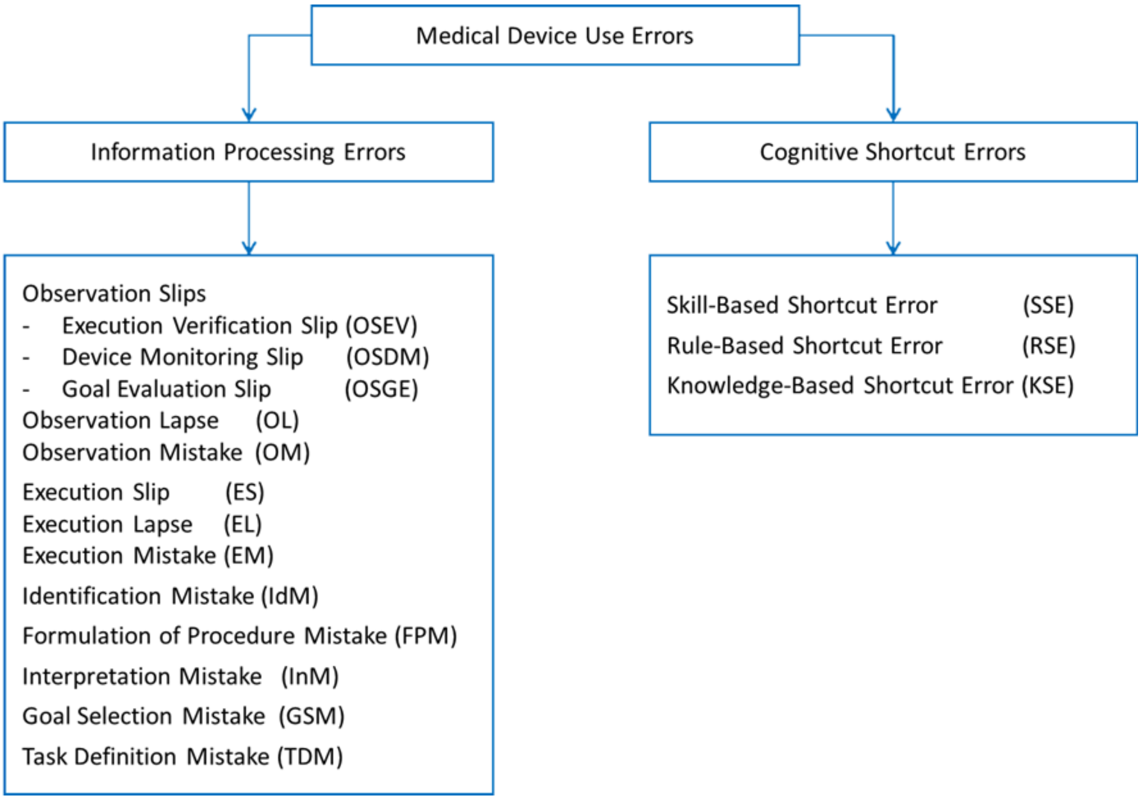

Medical Device Use Errors
Information Processing Errors
Cognitive Shortcut Errors
Observation Slips
- Execution Verification Slip (OSEV)
- Device Monitoring Slip (OSDM)
- Goal Evaluation Slip (OSGE)
Observation Lapse (OL)
Observation Mistake (OM)
Execution Slip (ES)
Execution Lapse (EL)
Execution Mistake (EM)
Identification Mistake (IdM)
Formulation of Procedure Mistake (FPM)
Interpretation Mistake (InM)
Goal Selection Mistake (GSM)
Task Definition Mistake (TDM)
Skill-Based Shortcut Error (SSE)
Rule-Based Shortcut Error (RSE)
Knowledge-Based Shortcut Error (KSE)

Figure 75. Use Error Taxonomy for interaction with medical devices, presented in **Section 5.1.2**. This taxonomy was systematically derived from two important DCMs, the Rasmussen's Decision-Ladder Framework [180] and Reason's Generic Error-Modelling System [199].

The future of ICS is quite difficult to predict, nevertheless as Monfort advised us, in his book *the Future* [4] about past visions of future technology, the futurology exercise is more convincing "…*when situated [framed] in existing, familiar experiences that we value and want to preserve.*" (p. 140). In this sense we should always envisioning future ICS as being used not only in leisure but also in working and training activities, and on both normal and safety-critical environments. When we consider complex new ICS, often the daVinci® surgical system (presented in **Section 4.1**) is given as an example of how the future landscape of a hospital will look like. In fact, we start seeing that the combination of both the challenges of IVEs – e.g, in the daVinci® surgeons rely on stereoscopic and AR images of the intervened region – and the challenges of operating using typical controls of a safety-critical medical systems is already a reality and it will be increasingly so with the advent of medical cyber-physical systems [234]. The same challenging combination starts to be proposed in other industries such as in the automotive one, where the car's manual interface starts to be accompanied by the displaying of virtual elements projected onto the windshield, in head-up displays, or in auto-stereoscopic central clusters [235].

Thus, in a time of renewed strength and innovation in the development of new interfaces, we are increasingly aware of the need to consider how well new ICS take into account the particularities of its users and its context of use. PCMs and DCMs will be in high demand and strategies to combine both types of user assessments in integrated applications may be the most suitable approach to deal with the upcoming complexity in ICS.

The work presented in this thesis was, for the most part, conducted at the Laboratory of Visualization and Perception, located at the Center for Computer Graphics, Guimarães, Portugal. The work described in Section 5.1, resulted from a 3 months internship at the Center for Devices and Radiological Health at the Food and Drugs Administration (Maryland, USA), under the supervision of Doctor Yi Zhang in collaboration with Doctor Paolo Masci.

This work has resulted in different scientific outputs, listed chronologically here:

2 Journal papers

Silva, C., Mendonça, C., Mouta, S., Silva, R., Campos, J., Santos, J. (2013) Depth cues and perceived audiovisual synchrony of biological motion. PLoS ONE, 8(11), e80096. DOI: 10.1371/journal.pone.0080096

Silva, C., Masci, P., Zhang, Y., Jones, P., Campos, J. (In Press). A Use Error Taxonomy for Improving Human-Machine Interface Design in Medical Devices. ACM SIGBED Review.

3 Conference Proceeding Papers

Silva, C. (2013). Audiovisual Perception in a Virtual World: An Application of Human-Computer Interaction Evaluation to the Development of Immersive Environments. *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM. DOI: 10.1145/2494603.2480340

Silva, C., Mouta, S., Santos J. (2016) Choosing audio devices on the basis of Listeners Spatial Perception: A case study of Headphones vs In-earphones. In the *IEEE 6th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Berlin, 2016, pp. 129-132. DOI: 10.1109/ICCE-Berlin.2016.7684737

Silva, C., & Campos, J. C. (2018). Towards a Simulation-Based Medical Education Platform for PVSio-Web. In the *IEEE 2018 International Conference on Graphics and Interaction (ICGI)* (pp. 1-8). DOI: 10.1109/ITCGI.2018.8602845

2 Conference presentations published as extended abstracts

Silva, C., Mouta, S., Basso, D., Santos, J., & Campos, J. (2015). Distance Perception in Immersive Environments-The Role of Photorealism. Presented at the 38th European Conference on Visual Perception (ECVP 2015, Liverpool). Perception (Vol. 44, pp. 321-321).

Silva, C., Mouta, S., Santos, J., Creissac, C. (2014). Spatial Limits for Audiovisual Unity assumption. Presented at the 37th European Conference on Visual Perception (ECVP 2014, Belgrade). Perception, vol. 43, 35 (ECVP Abstract supplement).

Other publications, in which the author of this one was not the main author, benefited from the contribution of the tasks carried out under the scope of this thesis:

J. A. Lamas, J., Silva, C., Silva, R., Mouta, S., Campos, J. C., & Santos (2015). Measuring end-to-end delay in real-time auralisation systems. in *10th European Congress and Exposition on Noise Control Engineering*, Maastrich, Netherderlands, 2015, pp. 791–796.

Finally, all the experiments' implementations are available at: https://github.com/Carlos-CCG/Thesis2019

References

[1] D. W. Maher John F Makowski, "Literary evidence for Roman arithmetic with fractions," 2001.

[2] P. E. Ceruzzi, *Computing : a concise history*. MIT Press, 2012.

[3] V. Bush, "As We May Think," *The Atlantic Monthly*, pp. 101–108, 1945.

[4] N. Montfort, *The Future*. Massachussetts: MIT Press, 2017.

[5] T. T. SIGCHI (Group : U.S.). Curriculum Development Group. *et al.*, *ACM SIGCHI curricula for human-computer interaction*. Association for Computing Machinery, 1992.

[6] A. Dix, "Human–computer interaction: A stable discipline, a nascent science, and the growth of the long tail," *Interact. Comput.*, vol. 22, no. 1, pp. 13–27, Jan. 2010.

[7] J. M. Carroll, "Conceptualizing a possible discipline of human–computer interaction," *Interact. Comput.*, vol. 22, no. 1, pp. 3–12, Jan. 2010.

[8] S. Reeves and Stuart, "Human-computer interaction as science," *Aarhus Ser. Hum. Centered Comput.*, vol. 1, no. 1, p. 12, Oct. 2015.

[9] Y. Rogers, "New theoretical approaches for HCI," 2004.

[10] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale, *Human-computer interaction*. Pearson/Prentice-Hall, 2004.

[11] I. S. MacKenzie, *Human-computer interaction.* Elsevier Science, 2012.

[12] J. Nielsen, *Usability engineering*. AP Professional, 1993.

[13] International Organization for Standardization, "ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems," 2010.

[14] Merriam-Webster, "Immersive | Definition of Immersive by Merriam-Webster," 2019. [Online]. Available: https://www.merriam-webster.com/dictionary/immersive. [Accessed: 13-Apr-2019].

[15] Merriam-Webster, "Virtual Reality | Definition of Virtual Reality by Merriam-Webster," 2019. [Online]. Available: https://www.merriam-webster.com/dictionary/virtual reality. [Accessed: 13-Apr-2019].

[16] K. Suzuki, S. Wakisaka, and N. Fujii, "Substitutional Reality System: A Novel Experimental Platform for Experiencing Alternative Reality," *Sci. Rep.*, vol. 2, no. 1, p. 459, Dec. 2012.

[17] A. Felnhofer *et al.*, "Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios," *Int. J. Hum. Comput. Stud.*, vol. 82, pp. 48–56, Oct. 2015.

[18] J.-H. T. Lin, "Fear in virtual reality (VR): Fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game," *Comput. Human Behav.*, vol. 72, pp. 350–361, Jul. 2017.

[19] H. T. Regenbrecht, T. W. Schubert, and F. Friedmann, “Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments,” *Int. J. Hum. Comput. Interact.*, vol. 10, no. 3, pp. 233–249, Sep. 1998.

[20] T. B. Sheridan, “Musings on Telepresence and Virtual Presence,” *Presence Teleoperators Virtual Environ.*, vol. 1, no. 1, pp. 120–126, Jan. 1992.

[21] G. Sziebig, “Achieving Total Immersion: Technology Trends behind Augmented Reality-A Survey,” in *Proceedings of the 9th WSEAS International Conference on Simulation, Modelling and Optimization*, 2009.

[22] W. Barfield, *Fundamentals of Wearable Computers and Augmented Reality, Second Edition*. CRC Press, 2015.

[23] A. S. Carlin, H. G. Hoffman, and S. Weghorst, “Virtual reality and tactile augmentation in the treatment of spider phobia: a case report,” *Behav. Res. Ther.*, vol. 35, no. 2, pp. 153–158, Feb. 1997.

[24] J. Jerald and Jason, *The VR book : human-centered design for virtual reality*. Association for Computing Machinery and Morgan & Claypool, 2015.

[25] M.-L. Ryan, *Narrative as virtual reality 2 : revisiting immersion and interactivity in literature and electronic media*, Parallax: Johns Hopkins University Press, 2015.

[26] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, “Surround-screen projection-based virtual reality,” in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’93*, 1993, pp. 135–142.

[27] R. West, M. J. Parola, A. R. Jaycen, and C. P. Lueg, “Embodied information behavior, mixed reality and big data,” 2015, vol. 9392, p. 93920E.

[28] A. Quigley and J. Grubert, “Perceptual and Social Challenges in Body Proximate Display Ecosystems,” in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI ’15*, 2015, pp. 1168–1174.

[29] I. E. Sutherland and I. E. Sutherland, “The Ultimate Display,” *Proc. IFIP Congr.*, vol. 2, pp. 506–508, 1965.

[30] J. E. Melzer and K. W. Moffitt, *Head-mounted displays : designing for the user*. [CreateSpace], 2012.

[31] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler, “Toward Low-Latency and Ultra-Reliable Virtual Reality,” *IEEE Netw.*, vol. 32, no. 2, pp. 78–84, Mar. 2018.

[32] J. A. Lamas, J., Silva, C., Silva, R., Mouta, S., Campos, J. C., & Santos, “Measuring end-to-end delay in real-time auralisation systems,” in *10th European Congress and Exposition on Noise Control Engineering, Maastrich, Netherderlands*, 2015, pp. 791–796.

[33] M. Di Luca, “New Method to Measure End-to-End Delay of Virtual Reality,” *Presence Teleoperators Virtual Environ.*, vol. 19, no. 6, pp. 569–584, Dec. 2010.

[34] S. Davis, K. Nesbitt, and E. Nalivaiko, “A Systematic Review of Cybersickness,” in *Proceedings of the 2014 Conference on Interactive Entertainment - IE2014*, 2014, pp. 1–9.

[35] J. Munafo, M. Diedrick, and T. A. Stoffregen, “The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects,” *Exp. Brain Res.*, vol. 235, no. 3, pp. 889–901, Mar. 2017.

[36] J. Reason, “Motion sickness: Some theoretical and practical considerations,” *Appl. Ergon.*, vol. 9, no. 3, pp. 163–167, Sep. 1978.

[37] F. Koslucher, E. Haaland, A. Malsch, J. Webeler, and T. A. Stoffregen, “Sex Differences in the Incidence of Motion Sickness Induced by Linear Visual Oscillation,” *Aerosp. Med. Hum. Perform.*, vol. 86, no. 9, pp. 787–793, Sep. 2015.

[38] J. Kim, C. Y. L. Chung, S. Nakamura, S. Palmisano, and S. K. Khuu, "The Oculus Rift: a cost-effective tool for studying visual-vestibular interactions in self-motion perception," *Front. Psychol.*, vol. 6, p. 248, Mar. 2015.

[39] S. Jones and S. Dawkins, "The Sensorama Revisited: Evaluating the Application of Multi-sensory Input on the Sense of Presence in 360-Degree Immersive Film in Virtual Reality," Springer, Cham, 2018, pp. 183–197.

[40] C. Loscos *et al.*, "The Museum of Pure Form: touching real statues in an immersive virtual museum," in *Proceedings of the 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 2004.

[41] C.-M. Wu, C.-W. Hsu, T.-K. Lee, and S. Smith, "A virtual reality keyboard with realistic haptic feedback in a fully immersive virtual environment," *Virtual Real.*, vol. 21, no. 1, pp. 19–29, Mar. 2017.

[42] R. Hinchet, V. Vechev, H. Shea, and O. Hilliges, "DextrES: Wearable Haptic Feedback for Grasping in VR via a Thin Form-Factor Electrostatic Brake," in *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, 2018, pp. 901–912.

[43] R. Sodhi, I. Poupyrev, M. Glisson, and A. Israr, "AIREAL," *ACM Trans. Graph.*, vol. 32, no. 4, p. 1, Jul. 2013.

[44] J. Vroomen and M. Keetels, "Perception of intersensory synchrony: A tutorial review," *Attention, Perception, Psychophys.*, vol. 72, no. 4, pp. 871–884, May 2010.

[45] H. Milgram P., Colquhoun, "A taxonomy of real and virtual world display integration," in *Mixed Reality: Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura, Eds. Springer-Verlag: Berlin, 1999, pp. 5–30.

[46] M. Kubovy, *The psychology of perspective and Renaissance art.* New York, NY, US: Cambridge University Press, 1986.

[47] "Blender Foundation — blender.org," 2019. [Online]. Available: https://www.blender.org/foundation/. [Accessed: 02-Apr-2019].

[48] C. Machover and S. E. Tice, "Virtual reality," *IEEE Comput. Graph. Appl.*, vol. 14, no. 1, pp. 15–16, Jan. 1994.

[49] V. Bruce, P. R. Green, and M. A. Georgeson, "Visual perception: Physiology, psychology, & ecology, 4th ed.," *Visual perception: Physiology, psychology, & ecology, 4th ed.* Psychology Press, New York, NY, US, p. xii, 483-xii, 483, 2003.

[50] E. B. Goldstein, *Sensation and perception*, 10th ed. Cengage Learning, 2016.

[51] H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation*," *Bell Syst. Tech. J.*, vol. 12, no. 4, pp. 377–430, Oct. 1933.

[52] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *Journal of Experimental Psychology*, vol. 47, no. 6. American Psychological Association, US, pp. 381–391, 1954.

[53] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses.," *J. Exp. Psychol.*, vol. 67, no. 2, pp. 103–112, 1964.

[54] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[55] A. Rohatgi, "WebPlotDigitizer - Extract data from plots, images, and maps," 2019. [Online]. Available: https://automeris.io/WebPlotDigitizer/. [Accessed: 02-Apr-2019].

[56] S. K. Card, Wi. K. English, and B. J. Burr, "Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT," *Ergonomics*, vol. 21, no. 8, pp. 601–613, Aug. 1978.

[57] I. S. MacKenzie and S. Jusoh, "An Evaluation of Two Input Devices for Remote Pointing," Springer,

Berlin, Heidelberg, 2001, pp. 235–250.

[58] I. S. MacKenzie, “A Note on the Information-Theoretic Basis for Fitts’ Law,” *J. Mot. Behav.*, vol. 21, no. 3, pp. 323–330, Sep. 1989.

[59] I. S. MacKenzie, “Fitts’ Law as a Research and Design Tool in Human-Computer Interaction,” *Human–Computer Interact.*, vol. 7, no. 1, pp. 91–139, Mar. 1992.

[60] I. S. MacKenzie and S. X. Zhang, “An empirical investigation of the novice experience with soft keyboards,” *Behav. Inf. Technol.*, vol. 20, no. 6, pp. 411–418, Jan. 2001.

[61] W. E. Hick, “On the Rate of Gain of Information,” *Q. J. Exp. Psychol.*, vol. 4, no. 1, pp. 11–26, Mar. 1952.

[62] R. Hyman, “Stimulus information as a determinant of reaction time.,” *Journal of Experimental Psychology*, vol. 45, no. 3. American Psychological Association, US, pp. 188–196, 1953.

[63] S. Seow, “Information Theoretic Models of HCI: A Comparison of the Hick-Hyman Law and Fitts’ Law,” *Human-Computer Interact.*, vol. 20, no. 3, pp. 315–352, Sep. 2005.

[64] T. K. Landauer, D. W. Nachbar, T. K. Landauer, and D. W. Nachbar, “Selection from alphabetic and numeric menu trees using a touch screen,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’85*, 1985, vol. 16, no. 4, pp. 73–78.

[65] A. Cockburn, C. Gutwin, and S. Greenberg, “A predictive model of menu performance,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’07*, 2007, p. 627.

[66] A. Ali and A. Liem, “The use of Formal Aesthetic Principles as a Tool for Design Conceptualisation and Detailing,” in *DS 81: Proceedings of NordDesign 2014, Espoo, Finland 27-29th August 2014*, M. Laakso and A. Aalto-Yliopisto, Eds. Aalto Design Factory, Aalto University, 2014, pp. 490–499.

[67] P. Hall, C. Heath, and L. Coles-Kemp, “Critical visualization: a case for rethinking how we visualize risk and security,” *J. Cybersecurity*, vol. 1, no. 1, p. tyv004, Dec. 2015.

[68] J. Ruiz, A. Bunt, and E. Lank, “A model of non-preferred hand mode switching,” in *Proceedings of Graphics Interface 2008*, 2008, pp. 49–56.

[69] S. K. Card, T. P. Moran, and A. Newell, “The Keystroke-Level Model for User Performance Time with Interactive Systems DOITree View project,” *Artic. Commun. ACM*, 1980.

[70] S. K. Card, *The Psychology of Human-Computer Interaction*. CRC Press, 1983.

[71] R. Bellamy, B. John, J. Richards, and J. Thomas, “Using CogTool to model programming tasks,” in *Evaluation and Usability of Programming Languages and Tools on - PLATEAU ’10*, 2010, pp. 1–6.

[72] S. Estes, “The Workload Curve,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 57, no. 7, pp. 1174–1187, Nov. 2015.

[73] L.-H. Teo, B. John, and M. Blackmon, “CogTool-Explorer,” in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*, 2012, p. 2479.

[74] P. Holleis, F. Otto, H. Hussmann, and A. Schmidt, “Keystroke-level model for advanced mobile phone interaction,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’07*, 2007, p. 1505.

[75] W. Gray, B. John, and M. Atwood, “Project Ernestine: Validating a GOMS Analysis for Predicting and Explaining Real-World Task Performance,” *Human-Computer Interact.*, vol. 8, no. 3, pp. 237–309, Sep. 1993.

[76] D. E. Kieras and D. E. Meyer, “An Overview of the EPIC Architecture for Cognition and Performance With Application to Human-Computer Interaction,” *Human–Computer Interact.*, vol. 12, no. 4, pp. 391–438, Dec. 1997.

[77] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, “An Integrated Theory of the Mind,” 2004.

[78] R. Ratcliff and J. J. Starns, “Modeling confidence and response time in recognition memory.,” *Psychol. Rev.*, vol. 116, no. 1, pp. 59–83, Jan. 2009.

[79] E. L. Nilsen, “Perceptual-Motor Control in Human-Computer Interaction.,” Michigan, 1996.

[80] J. R. Anderson, *How can the human mind occur in the physical universe?* Oxford: Oxford University Press, 2009.

[81] S. R. Ellis, D. R. Begault, and E. M. Wenzel, “Virtual Environments as Human-Computer Interfaces,” *Handb. Human-Computer Interact.*, pp. 163–201, Jan. 1997.

[82] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, “Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity,” in *Proceedings of the 1st Symposium on Applied perception in graphics and visualization - APGV ’04*, 2004, p. 39.

[83] J. J. Jerald, F. P. Brooks, M. C. Whitton, B. D. Adelstein, S. R. Ellis, and A. A. Lastra, “Scene-Motion-and Latency-Perception Thresholds for Head-Mounted Displays,” University of North Carolina, 2009.

[84] H. Fuchs and J. Ackerman, “Displays for augmented reality : Historical remarks and future prospects,” in *Mixed Reality-Merging Real and Virtual Worlds*, Ohm-sha, Springer-Verlag, 1999, pp. 1–11.

[85] O. Cakmakci and J. Rolland, “Head-Worn Displays: A Review,” *J. Disp. Technol.*, vol. 2, no. 3, pp. 199–216, Sep. 2006.

[86] R. Xiao and H. Benko, “Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 2016, pp. 1221–1232.

[87] J. L. Souman, P. R. Giordano, I. Frissen, A. De Luca, and M. O. Ernst, “Making virtual walking real,” *ACM Trans. Appl. Percept.*, vol. 7, no. 2, pp. 1–14, Feb. 2010.

[88] D. B. Kaber and T. Zhang, “Human Factors in Virtual Reality System Design for Mobility and Haptic Task Performance,” *Rev. Hum. Factors Ergon.*, vol. 7, no. 1, pp. 323–366, Sep. 2011.

[89] J. Andreano, K. Liang, L. Kong, D. Hubbard, B. K. Wiederhold, and M. D. Wiederhold, “Auditory Cues Increase the Hippocampal Response to Unimodal Virtual Reality,” *CyberPsychology Behav.*, vol. 12, no. 3, pp. 309–313, Jun. 2009.

[90] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. MÃ¼hlberger, “The impact of perception and presence on emotional reactions: a review of research in virtual reality,” *Front. Psychol.*, vol. 6, p. 26, Jan. 2015.

[91] A.-F. N. M. Perrin, H. Xu, E. Kroupi, M. Řeřábek, and T. Ebrahimi, “Multimodal Dataset for Assessment of Quality of Experience in Immersive Multimedia,” in *Proceedings of the 23rd ACM international conference on Multimedia - MM ’15*, 2015, pp. 1007–1010.

[92] C. Mendonça *et al.*, “Reflection orders and auditory distance,” in *Proceedings of Meetings on Acoustics*, 2013, vol. 19, no. 1, pp. 050041–050041.

[93] A. W. Bronkhorst and T. Houtgast, “Auditory distance perception in rooms,” *Nature*, vol. 397, no. 6719, pp. 517–520, Feb. 1999.

[94] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” 1978.

[95] E. A. Lehmann and A. M. Johansson, “Diffuse Reverberation Model for Efficient Image-Source Simulation of Room Impulse Responses,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, no. 6, pp. 1429–1439, Aug. 2010.

[96] J. E. Summers, “What exactly is meant by the term ‘auralization?,’” *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 697–697, Aug. 2008.

[97] R. B. Welch and D. H. Warren, “Immediate perceptual response to intersensory discrepancy.,” *Psychological Bulletin*, vol. 88, no. 3. American Psychological Association, US, pp. 638–667, 1980.

[98] G. Aschersleben, T. Bachmann, and J. Mu̇sseler, *Cognitive contributions to the perception of spatial and temporal events*. Elsevier, 1999.

[99] A. Vatakis and C. Spence, “Crossmodal binding: Evaluating the ‘unity assumption’ using audiovisual speech stimuli,” *Percept. Psychophys.*, vol. 69, no. 5, pp. 744–756, Jul. 2007.

[100] R. B. Welch, “Meaning, attention, and the ‘unity assumption’ in the intersensory bias of spatial and temporal perceptions,” *Adv. Psychol.*, vol. 129, pp. 371–387, Jan. 1999.

[101] P. Lennie, “The physiological basis of variations in visual latency,” *Vision Res.*, vol. 21, no. 6, pp. 815–824, Jan. 1981.

[102] J. H. Maunsell and J. R. Gibson, “Visual response latencies in striate cortex of the macaque monkey.,” *J. Neurophysiol.*, vol. 68, no. 4, pp. 1332–44, Oct. 1992.

[103] A. J. King and A. R. Palmer, “Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus,” *Exp. Brain Res.*, vol. 60, no. 3, pp. 492–500, Nov. 1985.

[104] W. Fujisaki, S. Shimojo, M. Kashino, and S. Nishida, “Recalibration of audiovisual simultaneity,” *Nat. Neurosci.*, vol. 7, no. 7, pp. 773–778, Jul. 2004.

[105] J. Vroomen, M. Keetels, B. de Gelder, and P. Bertelson, “Recalibration of temporal order perception by exposure to audio-visual asynchrony,” *Cogn. Brain Res.*, vol. 22, no. 1, pp. 32–35, Dec. 2004.

[106] D. Alais and S. Carlile, “Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 6, pp. 2244–7, Feb. 2005.

[107] R. Arrighi, D. Alais, and D. Burr, “Perceptual synchrony of audiovisual streams for natural and artificial motion sequences,” *J. Vis.*, vol. 6, no. 3, p. 6, Mar. 2006.

[108] N. F. Dixon and L. Spitz, “The Detection of Auditory Visual Desynchrony,” *Perception*, vol. 9, no. 6, pp. 719–721, Dec. 1980.

[109] A. Kopinska and L. R. Harris, “Simultaneity Constancy,” *Perception*, vol. 33, no. 9, pp. 1049–1060, Sep. 2004.

[110] Y. Sugita and Y. Suzuki, “Implicit estimation of sound-arrival time,” *Nature*, vol. 421, no. 6926, pp. 911–911, Feb. 2003.

[111] V. van Wassenhove, K. W. Grant, and D. Poeppel, “Temporal window of integration in auditory-visual speech perception,” *Neuropsychologia*, vol. 45, no. 3, pp. 598–607, Jan. 2007.

[112] S. A. Love, K. Petrini, A. Cheng, and F. E. Pollick, “A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments,” *PLoS One*, vol. 8, no. 1, p. e54798, Jan. 2013.

[113] R. Nijhawan and B. Khurana, Eds., *Space and Time in Perception and Action*. Cambridge: Cambridge University Press, 2010.

[114] M. Keetels and J. Vroomen, “The role of spatial disparity and hemifields in audio-visual temporal order judgments,” *Exp. Brain Res.*, vol. 167, no. 4, pp. 635–640, Dec. 2005.

[115] G. R. Engel and W. G. Dougherty, “Visual-Auditory Distance Constancy,” *Nature*, vol. 234, no. 5327, pp. 308–308, Dec. 1971.

[116] J. Lewald and R. Guski, “Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception,” *Neurosci. Lett.*, vol. 357, no. 2, pp. 119–122, Mar. 2004.

[117] D. H. Arnold, A. Johnston, and S. Nishida, “Timing sight and sound,” *Vision Res.*, vol. 45, no. 10, pp.

1275–1284, May 2005.

[118] K. Petrini, S. P. Holt, and F. Pollick, “Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception,” *J. Vis.*, vol. 10, no. 5, pp. 2–2, May 2010.

[119] A. Bierbaum, C. Just, P. Hartling, K. Meinert, A. Baker, and C. Cruz-Neira, “VR Juggler: a virtual platform for virtual reality application development,” in *Proceedings IEEE Virtual Reality*, 2001, pp. 89–96.

[120] M. M. Marcell, D. Borella, M. Greene, E. Kerr, and S. Rogers, “Confrontation Naming of Environmental Sounds,” *J. Clin. Exp. Neuropsychol.*, vol. 22, no. 6, pp. 830–864, Dec. 2000.

[121] J. Heron, D. Whitaker, P. V. McGraw, and K. V. Horoshenkov, “Adaptation minimizes distance-related audiovisual delays,” *J. Vis.*, vol. 7, no. 13, p. 5, Oct. 2007.

[122] V. Virsu, H. Oksanen-Hennah, A. Vedenpää, P. Jaatinen, and P. Lahti-Nuuttila, “Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations,” *Exp. Brain Res.*, vol. 186, no. 4, pp. 525–537, Apr. 2008.

[123] D. J. Lewkowicz, “Perception of auditory–visual temporal synchrony in human infants.,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 22, no. 5. American Psychological Association, US, pp. 1094–1106, 1996.

[124] P. Bertelson and M. Radeau, “Cross-modal bias and perceptual fusion with auditory-visual spatial discordance,” *Percept. Psychophys.*, vol. 29, no. 6, pp. 578–584, Nov. 1981.

[125] P. Bertelson, “Ventriloquism: Cross-modal patterning under spatial conflict,” in *cognitive contributions to the perception of spatial and temporal events.*, Amsterdam: Elsevier Science, 1999.

[126] W. D. Hairston, M. T. Wallace, J. W. Vaughan, B. E. Stein, J. L. Norris, and J. A. Schirillo, “Visual Localization Ability Influences Cross-Modal Bias,” *J. Cogn. Neurosci.*, vol. 15, no. 1, pp. 20–29, Jan. 2003.

[127] M. Kyto, K. Kusumoto, and P. Oittinen, “The Ventriloquist Effect in Augmented Reality,” in *2015 IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 49–53.

[128] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, 2001, pp. 99–102.

[129] F. Heller, A. Krämer, and J. Borchers, “Simplifying orientation measurement for mobile audio augmented reality applications,” in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*, 2014, pp. 615–624.

[130] D. Linares and J. López-Moliner, “quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups,” *R J.*, vol. 8, no. 1, pp. 122–131, 2016.

[131] P. D. Coleman, “An analysis of cues to auditory depth perception in free space.,” *Psychol. Bull.*, vol. 60, no. 3, pp. 302–315, 1963.

[132] International Organization for Standardization, “ISO 9613-1:1993 - Acoustics – Attenuation of sound during propagation outdoors – Part 1: Calculation of the absorption of sound by the atmosphere,” 1993. [Online]. Available: https://www.iso.org/standard/17426.html. [Accessed: 13-Apr-2019].

[133] T. Stoffregen, B. G. Bardy, L. J. Smart, and R. Pagulayan, “On the nature and evaluation of fidelity in virtual environments.,” *Virtual and adaptive environments: Applications, implications, and human performance issues.* Lawrence Erlbaum Associates Publishers, Stoffregen, Thomas: School of Kinesiology, University of Minnesota, 141A Mariucci Anena, 1901 4th St. S.E., Minneapolis, MN, US, 55414, pp. 111–128, 2003.

[134] G. Riccio, “Riccio, G. E. (1995). Coordination of postural control and vehicular control: Implications for multimodal perception and simulation of self-motion.,” in *Riccio, G. E. (1995). Coordination of postural control and vehicular control: Implications for multimodal perception and simulation of self-motion. Local*

applications of the ecological approach to human machine systems, 1995, pp. 122–181.

[135] Bing Wu, R. L. Klatzky, D. Shelton, and G. D. Stetten, “Psychophysical Evaluation of In-Situ Ultrasound Visualization,” *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 6, pp. 684–693, Nov. 2005.

[136] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker, “Distance perception in real and virtual environments,” *ACM Trans. Appl. Percept.*, vol. 2, no. 3, pp. 216–233, Jul. 2005.

[137] C. C. L. Silva and C. C.L., “Audiovisual perception in a virtual world,” in *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems - EICS ’13*, 2013, p. 175.

[138] S. Mouta, J. A. Santos, and J. Lopez-Moliner, “The time to passage of biological and complex motion,” *J. Vis.*, vol. 12, no. 2, pp. 21–21, Feb. 2012.

[139] T. Lokki, L. Savioja, R. Vaananen, J. Huopaniemi, and T. Takala, “Creating interactive virtual auditory environments,” *IEEE Comput. Graph. Appl.*, vol. 22, no. 4, pp. 49–57, Jul. 2002.

[140] T. Lokki and V. Pulkki, “Evaluation of Geometry-based Parametric Auralization,” in *Proceedings of the 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, 2002.

[141] C. Freksa, C. Habel, and K. F. Wender, Eds., *Spatial Cognition*, vol. 1404. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

[142] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, “Immersive virtual environment technology as a methodological tool for social psychology.,” *Psychol. Inq.*, vol. 13, no. 2, pp. 103–124, 2002.

[143] T. Iachini, Y. Coello, F. Frassinetti, and G. Ruggiero, “Body Space in Social Interactions: A Comparison of Reaching and Comfort Distance in Immersive Virtual Reality,” *PLoS One*, vol. 9, no. 11, p. e111511, Nov. 2014.

[144] T. L. Ooi and Z. J. He, “A distance judgment function based on space perception mechanisms: Revisiting Gilinsky’s (1951) equation.,” *Psychological Review*, vol. 114, no. 2. American Psychological Association, Ooi, Teng Leng: Department of Basic Sciences, Pennsylvania College of Optometry, 8360 Old York Road, Elkins Park, PA, US, 19027, tlooi@pco.edu, pp. 441–454, 2007.

[145] Z. Li, J. Phillips, and F. H. Durgin, “The underestimation of egocentric distance: evidence from frontal matching tasks,” *Attention, Perception, Psychophys.*, vol. 73, no. 7, pp. 2205–2217, Oct. 2011.

[146] M. V. Sanchez-Vives and M. Slater, “From presence to consciousness through virtual reality,” *Nat. Rev. Neurosci.*, vol. 6, no. 4, pp. 332–339, Apr. 2005.

[147] J. M. Loomis, “Three theories for reconciling the linearity of egocentric distance perception with distortion of shape on the ground plane.,” *Psychol. Neurosci.*, vol. 7, no. 3, pp. 245–251, 2014.

[148] M. Slater, P. Khanna, J. Mortensen, and I. Yu, “Visual Realism Enhances Realistic Response in an Immersive Virtual Environment,” *IEEE Comput. Graph. Appl.*, vol. 29, no. 3, pp. 76–84, May 2009.

[149] A. Higashiyama, “Horizontal and vertical distance perception: The discorded-orientation theory,” *Percept. Psychophys.*, vol. 58, no. 2, pp. 259–270, Mar. 1996.

[150] M. W. Scerbo and S. Dawson, “High Fidelity, High Performance?,” *Simul. Healthc. J. Soc. Simul. Healthc.*, vol. 2, no. 4, pp. 224–230, 2007.

[151] “www.makehumancommunity.org.” [Online]. Available: http://www.makehumancommunity.org/. [Accessed: 28-Apr-2019].

[152] P. Zimmons and A. Panter, “The influence of rendering quality on presence and task performance in a virtual environment,” in *IEEE Virtual Reality, 2003. Proceedings.*, pp. 293–294.

[153] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall, “Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments?,” *Presence Teleoperators Virtual Environ.*, vol. 13, no. 5, pp. 560–571, Oct. 2004.

[154] Insu Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual Realism Enhances Realistic Response in an Immersive Virtual Environment - Part 2," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 36–45, Nov. 2012.

[155] T. D. Parsons, A. A. Rizzo, C. G. Courtney, and M. E. Dawson, "Psychophysiology to Assess Impact of Varying Levels of Simulation Fidelity in a Threat Environment," *Adv. Human-Computer Interact.*, vol. 2012, pp. 1–9, 2012.

[156] J. M. Foley, N. P. Ribeiro-Filho, and J. A. Da Silva, "Visual perception of extent and the geometry of visual space," *Vision Res.*, vol. 44, no. 2, pp. 147–156, Jan. 2004.

[157] Se-Woon Jeon, Young-Cheol Park, and Dae Hee Youn, "Acoustic depth rendering for 3D multimedia applications," in *2012 IEEE International Conference on Consumer Electronics (ICCE)*, 2012, pp. 253–254.

[158] Sunmin Kim, Young Woo Lee, and Yoon Jae Lee, "3D sound rendering system based on relationship between stereoscopic image and stereo sound for 3DTV," in *2013 IEEE International Conference on Consumer Electronics (ICCE)*, 2013, pp. 324–325.

[159] T. Lee, Y. Baek, Y. Park, and D. H. Youn, "Stereo upmix-based binaural auralization for mobile devices," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 411–419, Aug. 2014.

[160] S. Poeschl, K. Wall, and N. Doering, "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence," in *2013 IEEE Virtual Reality (VR)*, 2013, pp. 129–130.

[161] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-An Overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, Nov. 1993.

[162] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. A. Santos, "On the Improvement of Localization Accuracy with Non-Individualized HRTF-Based Sounds," *J. Audio Eng. Soc.*, vol. 60, no. 10, pp. 821–830, Nov. 2012.

[163] C. Mendonça, G. Campos, P. Dias, and J. A. Santos, "Learning Auditory Space: Generalization and Long-Term Effects," *PLoS One*, vol. 8, no. 10, p. e77900, Oct. 2013.

[164] B. Gardner and K. Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone MIT Media Lab Perceptual Computing-Technical Report #280," 1994.

[165] A. Bierbaum, C. Just, P. Hartling, K. Meinert, A. Baker, and C. Cruz-Neira, "VR Juggler: a virtual platform for virtual reality application development," in *Proceedings IEEE Virtual Reality 2001*, pp. 89–96.

[166] Society of Automotive Engineers International., "ARP4754A: Guidelines for Development of Civil Aircraft and Systems - SAE International," 2010.

[167] M. Bozzano and A. Villafiorita, *Design and safety assessment of critical systems*. Auerbach Publications, 2011.

[168] M. R. Neuman *et al.*, "Advances in Medical Devices and Medical Electronics," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1537–1550, May 2012.

[169] K. W. Nam, J. Park, I. Y. Kim, and K. G. Kim, "Application of Stereo-Imaging Technology to Medical Field," *Healthc. Inform. Res.*, vol. 18, no. 3, p. 158, Sep. 2012.

[170] P. Chen, D., Mayer, E., Vale, J., Anstee, A., Mei, L., Darzi, A., Edwards, "A 3d stereo system to assist surgical treatment of prostate cancer," in *Workshop of Int Conf on Medical Image Computing and Computer Assisted Intervention*, 2008, pp. 9–17.

[171] L. Hatziharalambous, "An Introduction to Safety Critical Systems - White Paper," 2016.

[172] M. D. Harrison *et al.*, "Formal techniques in the safety analysis of software components of a new dialysis machine," *Sci. Comput. Program.*, vol. 175, pp. 17–34, Apr. 2019.

[173] M. R. Lyu and M. R., *Handbook of software reliability engineering*. IEEE Computer Society Press, 1996.

[174] N. G. Leveson and C. S. Turner, "An investigation of the Therac-25 accidents," *Computer (Long. Beach. Calif).*, vol. 26, no. 7, pp. 18–41, Jul. 1993.

[175] N. Leveson, "Medical Devices: the Therac-25," *Append. Safeware Syst. Saf. Comput.*, 1995.

[176] P. Masci, Y. Zhang, P. Jones, and J. C. Campos, "Extending STPA to Improve the Analysis of User Interface Software in Medical Devices," in *STAMP Workshop 2018*, 2018.

[177] AAMI, IEC, and ISO, "IEC 62366-1:2015 - Medical devices – Part 1: Application of usability engineering to medical devices," 2015.

[178] L. L. Leape and D. M. Berwick, "Five Years After To Err Is Human," *JAMA*, vol. 293, no. 19, p. 2384, May 2005.

[179] B. Middleton *et al.*, "Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA," *J. Am. Med. Informatics Assoc.*, vol. 20, no. e1, pp. e2–e8, Jun. 2013.

[180] J. Rasmussen and Jens, *Information processing and human-machine interaction : an approach to cognitive engineering*. North-Holland, 1986.

[181] P. Masci, "A preliminary hazard analysis for the GIP number entry software.," 2014.

[182] D. A. Norman, *The psychology of everyday things*, vol. 5. Basic books New York, 1988.

[183] J. Zhang, V. L. Patel, T. R. Johnson, and E. H. Shortliffe, "A cognitive taxonomy of medical errors," *J. Biomed. Inform.*, vol. 37, no. 3, pp. 193–204, Jun. 2004.

[184] J. Rasmussen, A. M. Pejtersen, and L. P. Goodstein, *Cognitive systems engineering*. Wiley, 1994.

[185] M. E. Hassall, P. M. Sanderson, and I. T. Cameron, "The Development and Testing of SAfER," *J. Cogn. Eng. Decis. Mak.*, vol. 8, no. 2, pp. 162–186, Jun. 2014.

[186] G. A. Miller, E. Galanter, and K. H. Pribram, *Plans and the structure of behavior.* New York: Henry Holt and Co, 1960.

[187] U. Neisser, *Cognition and reality: Principles and implications of cognitive psychology.* New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co, 1976.

[188] D. E. Human and R. Associates, "Human Error Understanding Human Behaviour and Error," 2016.

[189] P. Hollnagel, E., & Marsden, "Further Development of the Phenotype Genotype Classification Scheme for the Analysis of Human Erroneous Actions.," in *Inst. for Systems, Informatics and Safety.*, 1996.

[190] F. Gleeson and V. Hargaden, "Improving knowledge worker efficiency," in *2015 IEEE International Symposium on Technology and Society (ISTAS)*, 2015, pp. 1–8.

[191] N. Naikar, A. Moylan, and B. Pearce, "Analysing activity in complex systems with cognitive work analysis: concepts, guidelines and case study for control task analysis," *Theor. Issues Ergon. Sci.*, vol. 7, no. 4, pp. 371–394, Jul. 2006.

[192] F. Richters, J. M. Schraagen, and H. Heerkens, "Assessing the structure of non-routine decision processes in Airline Operations Control," *Ergonomics*, vol. 59, no. 3, pp. 380–392, Mar. 2016.

[193] N. A. Stanton and K. Bessell, "How a submarine returns to periscope depth: Analysing complex socio-technical systems using Cognitive Work Analysis," *Appl. Ergon.*, vol. 45, no. 1, pp. 110–125, Jan. 2014.

[194] D. P. Jenkins, N. A. Stanton, P. M. Salmon, G. H. Walker, and L. Rafferty, "Using the Decision-Ladder to Add a Formative Element to Naturalistic Decision-Making Research," *Int. J. Hum. Comput. Interact.*, vol. 26, no. 2–3, pp. 132–146, Mar. 2010.

[195] J. Taylor, *Human Error in Process Plant Design and Operations*. CRC Press, 2015.

[196] K. J. Vicente, *Cognitive Work Analysis*. CRC Press, 1999.

[197] J.-C. Le Coze, K. Pettersen, and T. Reiman, "The foundations of safety science," *Saf. Sci.*, vol. 67, pp. 1–5, Aug. 2014.

[198] N. G. Leveson, "Rasmussen's legacy: A paradigm change in engineering for safety," *Appl. Ergon.*, vol. 59, pp. 581–591, Mar. 2017.

[199] J. Reason, *Human error.* New York, NY, US: Cambridge University Press, 1990.

[200] I. A. Taib, A. S. McIntosh, C. Caponecchia, and M. T. Baysari, "A review of medical error taxonomies: A human factors perspective," *Saf. Sci.*, vol. 49, no. 5, pp. 607–615, Jun. 2011.

[201] J. M. Goldman, "Medical Devices and Medical Systems-Essential safety requirements for 5 equipment comprising the patient-centric integrated clinical environment (ICE)-Part 1: General requirements and conceptual model," 2008.

[202] International Organization for Standardization, "ISO 14971:2007 - Medical devices – Application of risk management to medical devices," 2007.

[203] F. Mason-Blakley, R. Habibi, J. Weber, and M. Price, "Assessing STAMP EMR with Electronic Medical Record Related Incident Reports: Case Study: Manufacturer and User Facility Device Experience Database," in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 2017, pp. 114–123.

[204] R. J. Mitchell, A. Williamson, and B. Molesworth, "Use of a human factors classification framework to identify causal factors for medication and medical device-related adverse clinical incidents," *Saf. Sci.*, vol. 79, pp. 163–174, Nov. 2015.

[205] D. A. Wiegmann and S. A. Shappell, "Human Error Analysis of Commercial Aviation Accidents Using the Human Factors Analysis and Classification System (HFACS)," United States. Office of Aviation Medicine, Feb. 2001.

[206] P. Elkin, M.-C. Beuscart-Zephir, S. Pelayo, V. Patel, and C. Nøhr, "The Usability-Error Ontology," in *Context sensitive health informatics : human and sociotechnical approaches*, M.-C. Beuscart-Zephir, M. Jaspers, C. Kuziemsky, C. Nøhr, and J. Aarts, Eds. Amsterdam: IOS Press BV, 2013, p. 211.

[207] M. B. Doumbouya, B. Kamsu-Foguem, H. Kenfack, and C. Foguem, "Argumentative reasoning and taxonomic analysis for the identification of medical errors," *Eng. Appl. Artif. Intell.*, vol. 46, pp. 166–179, Nov. 2015.

[208] S. Wiseman, P. Cairns, and A. Cox, "A taxonomy of number entry error," in *BCS-HCI'11 Proceedings of the 25th BCS Conference on Human-Computer Interaction*, 2011, pp. 187–196.

[209] Food and Drugs Administration, "MAUDE - Manufacturer and User Facility Device Experience." [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm. [Accessed: 21-Apr-2019].

[210] L. Lin, K. J. Vicente, and D. J. Doyle, "Patient Safety, Potential Adverse Drug Events, and Medical Device Design: A Human Factors Engineering Approach," *J. Biomed. Inform.*, vol. 34, no. 4, pp. 274–284, Aug. 2001.

[211] A. AAMI, "HE75: Human factors engineering – Design of medical devices," 2009.

[212] K. J. Vicente, K. Kada-Bekhaled, G. Hillel, A. Cassano, and B. A. Orser, "Programming errors contribute to death from patient-controlled analgesia: case report and estimate of probability," *Can. J. Anesth. Can. d'anesthésie*, vol. 50, no. 4, pp. 328–332, Apr. 2003.

[213] T. Walsh and P. C. W. Beatty, "Human factors error and patient monitoring," *Physiol. Meas.*, vol. 23, no. 3, p. 201, Aug. 2002.

[214] J. Brixey, T. R. Johnson, and J. Zhang, "Evaluating a medical error taxonomy.," *Proceedings. AMIA*

Symp., pp. 71–5, 2002.

[215] J. L. Martin, B. J. Norris, E. Murphy, and J. A. Crowe, “Medical device development: The challenge for ergonomics,” *Appl. Ergon.*, vol. 39, no. 3, pp. 271–283, May 2008.

[216] B. Kim *et al.*, “Safety-assured development of the GPCA infusion pump software,” in *Proceedings of the ninth ACM international conference on Embedded software - EMSOFT ’11*, 2011, p. 155.

[217] M. L. Bolton and E. J. Bass, “Generating Erroneous Human Behavior From Strategic Knowledge in Task Models and Evaluating Its Impact on System Safety With Model Checking,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 43, no. 6, pp. 1314–1327, Nov. 2013.

[218] M. D. Harrison, P. Masci, J. C. Campos, and P. Curzon, “Verification of User Interface Software: The Example of Use-Related Safety Requirements and Programmable Medical Devices,” *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 6, pp. 834–846, Dec. 2017.

[219] M. D. Harrison *et al.*, “Safety Analysis of Software Components of a Dialysis Machine Using Model Checking,” Springer, Cham, 2017, pp. 137–154.

[220] P. Masci, P. Oladimeji, Y. Zhang, P. Jones, P. Curzon, and H. Thimbleby, “PVSio-web 2.0: Joining PVS to HCI,” Springer, Cham, 2015, pp. 470–478.

[221] B. P. Kerfoot and N. Kissane, “The Use of Gamification to Boost Residents’ Engagement in Simulation Training,” *JAMA Surg.*, vol. 149, no. 11, p. 1208, Nov. 2014.

[222] S. R. Dawe *et al.*, “Systematic review of skills transfer after surgical simulation-based training,” *Br. J. Surg.*, vol. 101, no. 9, pp. 1063–1076, Aug. 2014.

[223] D. A. Cook *et al.*, “Technology-Enhanced Simulation for Health Professions Education,” *JAMA*, vol. 306, no. 9, pp. 978–988, Sep. 2011.

[224] B. Zendejas, A. T. Wang, R. Brydges, S. J. Hamstra, and D. A. Cook, “Cost: The missing outcome in simulation-based medical education research: A systematic review,” *Surgery*, vol. 153, no. 2, pp. 160–176, Feb. 2013.

[225] M. Ewertsson, M. Gustafsson, K. Blomberg, I. K. Holmström, and R. Allvin, “Use of technical skills and medical devices among new registered nurses: A questionnaire study,” *Nurse Educ. Today*, vol. 35, no. 12, pp. 1169–1174, Dec. 2015.

[226] S. Barry Issenberg, W. C. Mcgaghie, E. R. Petrusa, D. Lee Gordon, and R. J. Scalese, “Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review,” *Med. Teach.*, vol. 27, no. 1, pp. 10–28, Jan. 2005.

[227] C. M. Clancy and D. N. Tornberg, “TeamSTEPPS: Assuring Optimal Teamwork in Clinical Settings,” *Am. J. Med. Qual.*, vol. 22, no. 3, pp. 214–217, May 2007.

[228] J. H. Barsuk, W. C. McGaghie, E. R. Cohen, J. S. Balachandran, and D. B. Wayne, “Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit,” *J. Hosp. Med.*, vol. 4, no. 7, pp. 397–403, Sep. 2009.

[229] Y. Jun, K.-Y. Lee, K.-W. Gwak, and D. Lim, “Anatomic basis 3-D surgical simulation system for custom fit knee replacement,” *Int. J. Precis. Eng. Manuf.*, vol. 13, no. 5, pp. 709–715, May 2012.

[230] J. LeBlanc, C. Hutchison, Y. Hu, and T. Donnon, “A Comparison of Orthopaedic Resident Performance on Surgical Fixation of an Ulnar Fracture Using Virtual Reality and Synthetic Models,” *J. Bone Jt. Surgery-American Vol.*, vol. 95, no. 9, pp. e60-1–6, May 2013.

[231] J. Howe *et al.*, “Development of Virtual Simulations for Medical Team Training: An Evaluation of Key Features,” *Proc. Int. Symp. Hum. Factors Ergon. Heal. Care*, vol. 7, no. 1, pp. 261–266, Jun. 2018.

[232] “Awsomium,” *In-App Web Browser*. .

[233] N. Watson, S. Reeves, and P. Masci, “Integrating User Design and Formal Models within PVSio-Web,”

Nov. 2018.

[234] N. Dey, A. S. Ashour, F. Shi, S. J. Fong, and J. M. R. S. Tavares, “Medical cyber-physical systems: A survey,” *J. Med. Syst.*, vol. 42, no. 4, p. 74, Apr. 2018.

[235] N. Broy, F. Alt, S. Schneegass, and B. Pfleging, “3D Displays in Cars,” in *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI ’14*, 2014, pp. 1–9.

[236] K. Corti, G. Reddy, E. Choi, and A. Gillespie, “The Researcher as Experimental Subject: Using Self-Experimentation to Access Experiences, Understand Social Phenomena, and Stimulate Reflexivity,” *Integr. Psychol. Behav. Sci.*, vol. 49, no. 2, pp. 288–308, Jun. 2015.

[237] A. P. Field, J. Miles, and Z. Field, *Discovering statistics using R*. SAGE PublicationsSage UK: London, England, 2012.