John NA Brown, BA(Hons), BA, BEd, MSc

Unifying Interaction Across Distributed Controls In A Smart Environment

Using Anthropology-Based Computing To Make Human-Computer Interaction "Calm"

DISSERTATION

to gain the Joint Doctoral Degree
Doctor of Philosophy (PhD)
Alpen-Adria-Universität Klagenfurt
Fakultät für Technische Wissenschaften

in accordance with
The Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments
Alpen-Adria-Universität Klagenfurt | Universitat Politècnica de Catalunya
Università degli Studi di Genova

First Supervisor
Univ.-Prof. Dr. Martin Hitz
Institut für Informatik-Systeme
Alpen-Adria-Universität Klagenfurt, Austria

Second Supervisor
Prof. Dr. Andreu Català Mallofré
Centre Tecnològic de Recerca per a la Depencia i la Vida Autònoma
Universitat Politècnica de Catalunya, Spain

February 3rd, 2014

Acknowledgments

This PhD Thesis has been developed in the framework of, and according to, the rules of the Erasmus Mundus Joint Doctorate on Interactive and Cognitive Environments EMJD ICE [FPA n° 2010-0012] with the cooperation of the following Universities:

Alpen-Adria-Universität Klagenfurt – AAU

Queen Mary, University of London – QML

Technische Universiteit Eindhoven – TU/e

Università degli Studi di Genova – UNIGE

Universitat Politècnica Catalunya – UPC

According to ICE regulations, the Italian PhD title has also been awarded by the Università degli Studi di Genova.

This work was supported by Lakeside Labs GmbH, Klagenfurt, Austria with funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant 20214/21530/32602.

Lakeside Labs GmbH – Klagenfurt, Austria

First Reviewer
Univ.-Prof. Dr. Martin Hitz
Institut für Informatik-Systeme
Alpen-Adria-Universität Klagenfurt, Austria

Second Reviewer
Prof. Dr. Andreu Català Mallofré
Centre Tecnològic de Recerca per a la Depencia i la Vida Autònoma
Universitat Politècnica de Catalunya, Spain

Declaration of Honor

I hereby confirm on my honor that I personally prepared the present academic work and carried out myself the activities directly involved with it. I also confirm that I have used no resources other than those declared. All formulations and concepts adopted literally or in their essential content from printed, unprinted or Internet sources have been cited according to the rules for academic work and identified by means of footnotes or other precise indications of source.

The support provided during the work, including significant assistance from my supervisor has been indicated in full.

The academic work has not been submitted to any other examination authority. The work is submitted in printed and electronic form. I confirm that the content of the digital version is completely identical to that of the printed version.

I am aware that a false declaration will have legal consequences.

John NA Brown Klagenfurt, February 3rd, 2014

Abstract

Rather than adapt human behavior to suit a life surrounded by computerized systems, is it possible to adapt the systems to suit humans? Mark Weiser called for this fundamental change to the design and engineering of computer systems nearly twenty years ago. We believe it is possible and offer a series of related theoretical developments and practical experiments designed in an attempt to build a system that can meet his challenge without resorting to black box design principles or Wizard of Oz protocols. This culminated in a trial involving 32 participants, each of whom used two different multimodal interactive techniques, based on our novel interaction paradigm, to intuitively control nine distributed devices in a smart home setting. The theoretical work and practical developments have led to our proposal of seven contributions to the state of the art.

Acknowledgements

Professor Doctor Hitz and Professor Doctor Català have had a permanent effect on my understanding of what it means to be a professor and a mentor. To them both I offer my sincere thanks. Gentlemen, I will try to apply all that you have taught me.

Throughout this document, credit is given to my co-authors. Saskia Bayerl showed me how to better explain complex ideas, Hamid Bouchachia challenged me to open a black box, and István Fehévari built the initial prototypes just because it seemed like I had a good idea.

Four other co-authors started as my students: Franz Huber, Karl-Heinz Pirolt, Florian Bacher and Christophe Sourisse did a fine job of collecting data, reviewing literature, and drafting arguments.

I must make special mention of Dr. Gerhard Leitner who is imparting some of his own sensitivity and wisdom to smart homes, and was kind enough to share it with me, too. Mr. Anton Fercher built, coded, designed and/or improved every piece of technology used in my final experiment and helped to run the trials, too. I would have been lost without him.

Above all other co-authors listed here, Bonifaz Kaufmann has been my brother-in-arms.

Lady and gentlemen, without your contributions, this dissertation would consist solely of a series of untried ideas, and my time here would have been much less pleasant.

There are others who have made my work possible even though they did not directly collaborate on my publications or experiments: without the generous and insightful help of Uwe Dutschmann, this dissertation would be pencil on paper, and I would have been a lot less happy during the process. International scientist Vera Mersheeva has become my new sister, and I would not have completed this work without her support.

The research teams at all five of the ICE universities have provided valuable feedback and suggestions in a collegial atmosphere. My deepest thanks go to the ICE fellows and other PhD candidates, the post docs, professors, researchers, technicians and assistants with whom I have worked and played at AAU, UPC, TU/e, QMUL, and UNIGE. Gentlemen and ladies, you have been my colleagues, my collocutors, and my collaborators. More than that; you have been my friends. I hope the last of those, at least, will continue.

The professors, administrators, and assistants who managed the ICE program during my tenure deserve more and better recognition than I can show here.

Dedication

For my family and friends around the world, most of whom will never read this.

This dissertation and the three years of work that have gone into it could never have been completed without your support, assistance, and kindness. I will try to show my gratitude by paying it forward for the rest of my life.

John

Publications from this Thesis Work

Journal

1) Gerhard Leitner, Martin Hitz, Anton Josef Fercher, John NA Brown. "Aspekte der HCI im Smarthome". In: HMD – Praxis der Wirtschaftsinformatik – Human Computer Interaction. Heft 294 (Dezember 2013), dpunkt-Verlag.

Monograph and Keynote Lecture

2) Brown, John NA. "It's as Easy as ABC: Introducing Anthropology-Based Computing" In Advances in Computational Intelligence, pp. 1-16. Springer Berlin Heidelberg, 2013.

Extended Abstract and Plenary Session Presentation

3) Brown, John NA. "Expert Talk for Time Machine Session: Designing Calm Technology "… as Refreshing as Taking a Walk in the Woods"," 2012 IEEE International Conference on Multimedia and Expo, vol. 1, pp. 423, 2012.

Full Papers and Conference Presentations

4) Brown, John NA, Bonifaz Kaufmann, Franz J. Huber, Karl-Heinz Pirolt, and Martin Hitz. ""… Language in Their Very Gesture" First Steps towards Calm Smart Home Input." In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pp. 256-264. Springer Berlin Heidelberg, 2013.

5) Brown, John NA, Bonifaz Kaufmann, Florian Bacher, Christophe Sourisse, and Martin Hitz. "" Oh, I Say, Jeeves!" A Calm Approach to Smart Home Input." In Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, pp. 265-274. Springer Berlin Heidelberg, 2013.

Workshop Papers and Presentations

6) Brown, J. N. A., Gerhard Leitner, Martin Hitz and Andreu Català Mallofré. (2014, April). A Model of Calm HCI. In S Bakker, D Hausen, T Selker, E van den Hoven, A Butz, B Eggen (Editors) Peripheral Interaction: Shaping the Research and Design Space. Workshop at CHI2014, Toronto, Canada. ISSN: 1862-5207.

7) Brown, J. N. A., P. S. Bayerl, Anton Fercher, Gerhard Leitner, Andreu Català Mallofré, and Martin Hitz. (2014, April). A Measure of Calm. In S Bakker, D Hausen, T Selker, E van den Hoven, A Butz, B Eggen (Editors) Peripheral Interaction: Shaping the Research and Design Space. Workshop at CHI2014, Toronto, Canada. ISSN: 1862-5207.

"Ich glaube, die Herren Rezensenten engagierter und nicht engagierter Art werden wieder einige Anwandlungen von Drehkrankheit bekommen, dagegen werden Freunde eines gesunden Spaßes die Spaziergänge, die ich ihnen da bereite, sehr amüsant finden. Das Ganze ist leider wieder von dem schon so übel beleumdeten Geiste meines Humors angekränkelt, und finden sich auch oft Gelegenheiten, meiner Neigung zu wüstem Lärm nachzugehen."

Gustav Mahler,
criticising his own work in a letter to Bruno Walter, 1878
reprinted in Blaukopf, ed., Gustav Mahler Briefe, 191; Martner, ed., Selected Letters,189

"El caminant, quan entra en aquest lloc,
comença a caminar-hi a poc a poc;
compta els seus passos en la gran quietud:
s'atura, i no sent res, i està perdut."

Joan Maragall,
La fageda d'en Jordà, Seqüències, 1911

"So many people today — and even professional scientists — seem to me like someone who has seen thousands of trees but has never seen a forest. A knowledge of the historic and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering. This independence created by philosophical insight is — in my opinion — the mark of distinction between a mere artisan or specialist and a real seeker after truth."

Albert Einstein,
from a letter to R.A. Thorton, 1944
EA-674, Einstein Archive, Hebrew University, Jerusalem

Contents

CHAPTER 1

Introduction

The evolution and adaptation of humans is intractably intertwined with the evolution and adaptation of our technology [144]. This was true when we added wooden handles to stone adzes, and it is true today. Nineteen years ago Mark Weiser warned that Ubiquitous Computing (UC) would require the development of a tremendous change to the way in which we interact with computers and the machines that house them [154]. His prediction of ubiquity has come true and we are surrounded by computerized systems that shape most of our day-to-day interactions with technology [155]. Despite the ubiquity of computers and computerization we have done very little to adapt the proliferating technology to our new way of life [152]. Weiser's proposed solution, "Calm Technology" (CT) describes tools made to suit the natural ways in which humans perceive, process, and respond to the world [153]. He called for a re-imagining of how we interact with computerized systems, so that the entire process could become more suitable to human perceptual abilities and limitations [156].

Bardzell and Bardzell have suggested that Weiser's prediction of UC was descriptive of a technological trend that was already in the process of being met, while his call for CT was more theoretical or philosophical [15]. In an attempt to bring the two sides together, we propose a new, human-centered approach to Human-Computer Interaction (HCI) called "Anthropology-Based Computing" (ABC) [23].

ABC replaces the standard machine-like concept of a human in previous models of interaction [125, 2, 49] with the three-tiered Brown-Hitz model of "Calm" interaction that incorporates the natural *reflexive*, *pre-attentive* and *attentive* levels of interaction that shape the way humans interact with the world [27].

The richer understanding of HCI afforded by this new model leads us in two directions. The first is the attempt to develop a prototypical quantitative measure of Weiser's "Calm" [28], with the intent that such a measure should help to establish "Calm" as a formal aspect of technological design [68] and may, eventually lead to the mitigation of techno-stress, ergonomic issues, and the incidents and accidents that result from interaction with computerized tools and devices that are not truly human-centered [136, 135].

The second is the development of tools that would practically demonstrate the concept of "ABC" [22]. As a result, we have created a toolset that enables an untrained person to interact intuitively with a subset of the heterogeneous computerized devices found in the standard home of the 21st Century.

This toolset and the associated mental models have been tested with 32 un-familiarized volunteers intuitively performing a variety of normal day-to-day tasks in two very different multimodal manners, one centering on natural speech and the other centered on familiar gestures.

1.1 Technology Usually Becomes "Calm" Over Time.

2.34 million years ago, small figures sat on the edge of a short cliff in what is now called Kenya, and made stone axes or adzes [137]. Over two thousand pieces of evidence define this workshop as the site of the earliest known tool-makers in history [115]. These makers, pre-human anthropoids of the species Homo habilis were experts at precision work, and the tools they produced are the precedents of every hammer and axe that has been made since. In fact, it is not an exaggeration to say that, if this was the beginning of tool making, then all of the technological advances that have happened since can trace their roots to this one site in the Rift Valley [144]. Stone axes were in the hands of our ancestors for another 2.3 million years before the early humans of the Mesolithic period took the next major step and improved not only efficiency, but also user safety and comfort by adding handles [42]. Axes continued to evolve and became one of the most common tools in the world [130].

It is safe to say, then, that humans have used axes for a long time. When in *flow* [51], an expert uses an axe without conscious thought or attention [108]. That is to say that the conscious part their brain is not engaged [12]. Instead, they engage in a combination of reflexive and pre-attentive interactions [87]. The user makes micro-adjustments to their grip according to haptic feedback during the swing, and reads the reactive forces as an unconscious guide for the next swing. The axe itself, however, only becomes the focus of attention if something unexpected happens [11]. Otherwise, the user's focus is on the task rather than the tool, and they are free to perform that task while being peripherally attentive to other environmental stimuli [14]. It seems that axes are an example of "Calm Technology" [22].

Sometime in the last three decades, computer interface devices replaced hammers and axes as the most ubiquitous tools [3], but these new devices have not had enough time to evolve into a "Calm Technology" [23]. Computer interaction technology fails to be "Calm" in several ways. The tool or the system that underlies it can still present interruptive signals and error messages at any moment, demanding our attention and distracting us from our work or play to deal with technical issues.

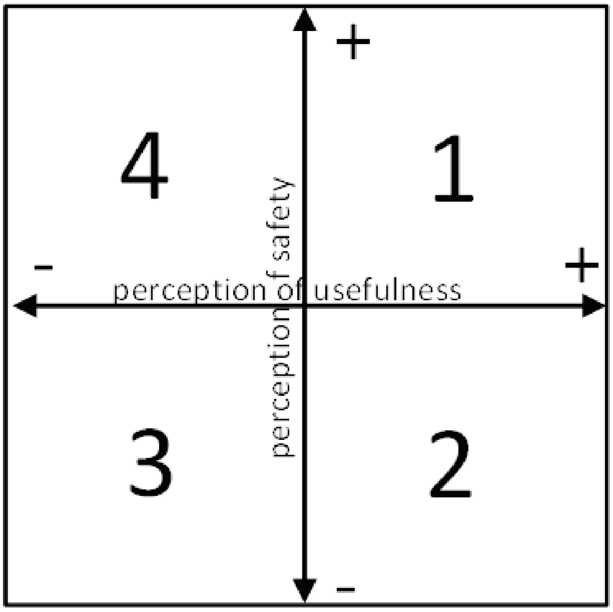


Figure 1: Perception of Technology Safety & Usefulness

Figure 1 illustrates the overlapping realms of perceived tool safety and perceived tool usefulness, for the purpose of showing how tools evolve. Let us define tool usefulness as the degree to which the tool can be applied successfully to day-to-day tasks, and tool safety as the degree to which a tool can be used as intended without injuring the user. It would be intellectually challenging and informing to populate this graphic with precise data for specific technology; showing when a tool went, for example, from less safe and less useful (quadrant 3) to more safe and more useful (quadrant 1). Our purpose is to use the graph simply to illustrate that technology moves in our perception.

When the first hand axes were made, as discussed earlier, they were instantly somewhat useful, but they were not very safe to use. This would place them in the lower left of quadrant 2. The technology spread across continents before it became significantly safer with the addition of handles. This addition moved the tool from towards the center of quadrant 1. As a truly ubiquitous tool for thousands of years, the axe eventually settled somewhere in the upper right-hand corner of quadrant 1. With other changes in technology and in our lives, the axe has come to be perceived as less useful. It may have dropped into quadrant 4 or, in the perception of the part of the population that has never learned to use an axe, even to quadrant 3. Again, precise times and associated Cartesian coordinates are not the point here. We are simply illustrating the movement across these quadrants of a technology. One could do the same with any other technology, be it simple or complex. Electricity could be mapped, showing the progress it made from being perceived as a dangerous toy (quadrant 3) [kite] to a completely unchallenged part of our day-to-day lives (quadrant 1) [22].

Many factors affect the movement of a tool or technology from one quadrant to another, including material resources, societal pressures, and related technological advances [144]. The point in raising this is to show that while the progress of axes could be mapped for literally millions of years, and electricity for hundreds, computers have simply not had the time to achieve a reliable position. Weiser's assertion of the need for CT is based on the idea that the computer has come to be perceived as so useful that we have put it to work all around us without taking the time to make it safe.

Consider the early computer. It was certainly safe (unlikely to harm the user during normal use), it just wasn't particularly useful (quadrant 4). Even as it became more useful, eventually moving from Weiser and Brown's Era of Mainframe computing to Personal Computing, the truth about the biomechanical dangers of using computers for both work and play were only beginning to be understood [21]. As a result, computers were already perceived to be harmless but very useful (quadrant 1) before we began to associate injuries to certain specific aspects of HCI, such as the use of the mouse [1] and the keyboard [84, 141], or the monitor, the desk, and the chair [30].

It is interesting to note that the cockpit of all planes above a minimum size must be equipped with an axe for use in case of an emergency. This is covered in Canada under Civil Aviation Regulation (CAR) 705.92, and in the United States under Federal Aviation Regulation (FAR) 121.309. It is understood that cabin crew will only use the axe if necessary. Cabin crews are not trusted to make similar decisions regarding personal computers, whether they are laptops, palmtops, tablets or smartphones. The bans on flying a commercial aircraft while using a PC, or on texting while driving, show that our perception of computers has changed from harmless to harmful. The computer is moving from quadrant 1 to quadrant 2. While the axe and electricity progressed from bottom to top of our graph, computers have been moving from top to bottom. We propose that this movement can be reversed if the HCI community embraces the concept of making deliberate, drastic evolutionary changes.

It might seem unreasonable to expect to make drastic evolutionary improvements to the use of a tool in a single generation, but that is what Weiser called for [153]. After predicting the advent of Ubiquitous Computing [155], he stated flatly that Calm Technology would be necessary to help us cope safely with a computerized environment [152]. Many researchers have tried to pick up Weiser's mantle and find the path to "Calm". Norbert Streitz carried on along Weiser's path towards the disappearing computer [145]. Hiroshi Ishii sought to create tangible input and output devices that were conceptually removed from direct computer control [92, 93]. Saskia Bakker focused on the principle of peripheral interaction as a requirement of "Calm" and developed tools for use by non-experts during engagement in other tasks [12, 13, 14]. With few exceptions, though, the concept of "Calm Technology" has fallen out of focus [15]. The abandonment of Weiser's theory was exemplified in Rogers' call to reignite some excitement (as opposed to calm) among computer users [137]. Progress towards the ubiquity of distributed and embedded systems continues at a growing pace [46] with little or no concern for the fact that such systems contribute to psychological [85] and physical trauma [11] as well as incidents and accidents in high-risk industries where overlapping, heterogeneous computerized control systems have become the norm [135].

1.2 Heterogeneous Designs Amplify the Problem

To date, most smart environments are a cluster of heterogeneous, intentionally incompatible subsystems [146]. No longer only true for nuclear power plants, aircraft, and space stations, this is now also true for smart homes. It has been suggested that the ubiquity of smartphones and other similar devices has surrounded individuals in smart environments which they carry with them [116]. Though nearly universal [99], this practice could have dire repercussions because using a phone while driving adds a layer of complexity to the tasks performed [59] and also adds a potential need for concentration when attention should be focused on the road ahead [11]. In fact, some would argue that using a phone while walking exemplifies the same problem but with a substantial reduction of mass and velocity imparting a lesser demand on reaction times and a lesser risk of injury or death.

It is clear that adding demanding tasks increases risk of failure [119]. The heterogeneity of task performance magnifies this risk by creating a constant demand for readiness to devote full attention to more than one cognitively-different complex interface [89]. Use of technology without a sense of being in control generates stress [18], exactly as Weiser predicted [152]. Adding the threat of negative outcomes amplifies that stress [136]. It is bad enough that these risks are accepted in some high-risk industries [121]. With the proliferation of deliberately smart environments like smart homes [39], as well as the incidentally-smart environment created by mobile computing, these and other Human factors issues also seem likely to move from industry into day-to-day life.

The fundamental problem of implementing "Calm" methods of interfacing with computerized tools is the lack of a unifying theoretical basis that would allow one to manifest the ephemeral, qualitative value of Weiser's "Calm" with the practical, technical, quantitative values required for a device to become "ubiquitous" [15]. The intent of this thesis is to propose modifications to, or departures from, the current state of the art of both the idea of CT and the practice of UC in order to achieve the original high-level concept of enhanced living through computer-based technology that has been woven into the patterns of everyday life and so, made to fit the human environment.

It is hoped that the right innovations would allow for seamless transition across modalities and devices, and for perception of a smart environment as a single, holistic entity. In order to achieve measurable demonstration of real world application, it was first necessary to develop theories that would inform the construction of the required hardware and software. We will introduce our additions to the theory of HCI in Chapter 3, and our technological contributions in Chapter 4. By applying these theories and prototypes we aim to demonstrate a paradigm of interaction that will allow the computer user to attend to other matters peripherally, as safely as though they were using an axe.

1.3 Formal Research Questions

Problems: Smart environments require the use of many heterogeneous devices. Using more than one device at a time is difficult in a way that cannot yet be measured, and the difficulty seems to increase if the interactive techniques are heterogeneous.

H0: Interaction with the distributed devices of a smart environment cannot be conceptually unified so that it can be perceived and performed as though with a single "calm" system.

Research Questions:

Q1) Can Weiser's "Calm" be defined in scientific manner?

Q2) Can Weiser's "Calm", as defined, be used as a quantitative metric?

Q3) Can natural human peripheral interaction be applied to smart home controls?

Q4) Can a smart home ontology be customized dynamically to suit each user?

Q5) Can interaction be designed so that users perceive many distributed, heterogeneous devices as a single, holistic, "calm" system?

The process of attempting to answer these five questions has led to the development of seven innovations. We list them here as contributions to the state of the art. We will address these contributions again with a point-by-point review of our accomplishments in Chapter 6.1.

1.4 Contributions

As mentioned above, there is a rift between the realization of Weiser's practical UC and his more theoretical CT. Each contribution described in this dissertation is a building block in a bridge that spans that rift. Before discussing these contributions in detail, we offer Figure 2 in which the contributions are presented in clockwise order. First it was necessary to develop a new theory of HCI in general, and a new model of HCI in particular, that could provide a scientific basis for CT (1). The next step was the development of strategies regarding the implementation of the theories, enabling people to interact in a natural and intuitive manner (2, 3 and 4). These strategies were then tested: first as a simple multimodal input device for using intuitive gestures in game play (5), and then in the fully-developed context of interacting with distributed interfaces in a smart environment (6). Our final contribution is a prototypical metric for the quantitative evaluation of "Calm" (7).

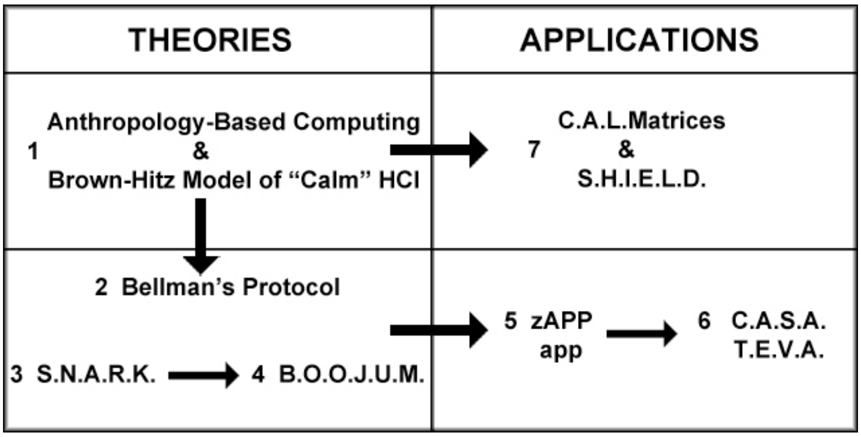


Figure 2: Theoretical and applied contributions

1) Addressing Q1, we propose the Brown-Hitz model of “Calm” Human-Computer Interaction, based on the understanding of how humans perceive, process, and react to stimuli in the real world. This understanding is the root of the truly human-centered interaction that we call “Anthropology-Based Computing” (ABC).

   In order to design an interactive system that is in line with Weiser’s concept of “Calm Technology”, we must first define “Calm”. In order to unite theory and practice, our definition must have practical value. Previous models of HCI have presented the entities on either side of the interface as analogs of one another [125]. Like the machine, the human receives input, processes it, and then generates output. We propose a new model that reflects the fact that humans can take in, process, and even react to information in a myriad of different ways, using many different parts of our central nervous and cerebral systems. We propose a generalization into a three-tiered model, identifying and differentiating between the “reflexive”, “pre-attentive” and “attentive” processes. This will be discussed in greater detail in Chapter 3.

2) Addressing Q5, we propose the Bellman’s Protocol for smart home interaction.

   Weiser’s requirement that machines “...fit the human environment instead of forcing humans to enter theirs” [155] requires a new paradigm for designers, engineers and all those who will live or work in the new environment. The problem is one of mental mapping, navigation, and control.

   In the poem, “The Hunting of the Snark” by Oxford mathematician C. L. Dodson [34], the Bellman is captain of a ship who navigates without a map by using a communication tool that was ubiquitous at the time of writing (a ship’s bell). It was well understood that ringing a bell could not allow one to navigate through the real world, and it was understood that one needed an accurate map in order to find a path to one’s intended destination. Against all of the pervading logic of the time, the Bellman simply rings his bell and arrives at his goal without worrying about navigating at all.

   That is the model for our Bellman’s Protocol: an interaction method that should seem too simple to be possible. The simplicity is achieved through hard, detailed advance work by the designer, naturalistic conscious and unconscious human input, and a naturalistic, *humanesque* output. The concept is explained in detail in Chapters 3, 4 and 5.

3) Addressing Q3, we propose a system for combining multiple human outputs into a more probable computer input, the S.N.A.R.K..

Peripheral interaction is a necessary component of intuitive human communication [83]. Peripheral interaction with computers, however, often provides only subtle signals [12] that are difficult to detect under laboratory conditions and almost impossible to detect with any accuracy in the real world [13]. As a result, many previous attempts at detecting peripheral human input have largely been abandoned or relegated to "black box" design space. At best, attempts to capture these signals without extensive environmental equipment result in many false positives and negatives [73].

In the same poem cited above, the Bellman and his crew are hunting for a Snark, a legendary creature that no one has seen before, but that all of them believe it may be possible to find. They all want to find the creature, and the crew has been assembled based on a wide variety of skills and backgrounds, but they do not know how to proceed.

The impossible nature of this quest inspires the name of the method by which natural human behavior, both conscious and unconscious, can be used as computer input. The S.N.A.R.K. works as a triple-redundancy failsafe, on the model of the software developed for machine-machine communication in old satellite systems [95], by Synchronizing Natural Actions and Reacting Knowledgeably. Earlier versions of the S.N.A.R.K. were tested and reported in [25] and [26]. The concept will be discussed in detail in Chapter 4.

4) Addressing Q4, we propose an open ontology that accommodates new devices and changing environments, as well as changing users and their individual preferences, the B.O.O.J.U.M..

To date, the ontologies that underlie smart environments are proprietary and designed to suit a specific realm of associated devices [37, 45, 48, 55, 75,]. New devices cannot be added without customization [52, 104]. Furthermore, a different sort of customization allows a personal experience for the user [32, 38], but not for more than one user [33], and not in a manner that is easily adaptable [40, 57, 58, 127]. Generally, the systems are standardized [7] and require that the user(s) adapt themselves to their environment [9, 54], in direct opposition to Weiser's intent [25, 26].

In the final reference to the poem cited above, we note that the end of the story reveals that the Snark is really a Boojum; a creature that is something different to everyone who imagines it. What's more, a Boojum is invisible and goes undetected.

This seemed the perfect name for a smart home ontology based on allowing each individual user to use their own preferred commands and names for each device. The B.O.O.J.U.M. (Brown's Open Ontology for Joint User Management) is made up of two parts. The first is a device ontology based on formally recording the common values in each user's mental map of their home. This does not require any changes to the underlying ontology of the networked and embedded systems in the home; it is intended to assist the users, not the software. The second part of the B.O.O.J.U.M. is a 20-item command lexicon derived from a survey of 435 atomic use cases based on activities of daily living. This is further described in Chapter 4.

5) Addressing Q3, we present a means of using a smartphone as an input device for capturing gestures based on pre-existing mental models: the zAPP App.

   It is "common knowledge" that gestures designed for HCI are difficult to learn [91, 96]. Despite that, gestures are a fundamental part of normal human communication [80]. That may be why the development of gesture-based interaction tools continues [43, 44, 69, 86, 90, 94, 97, 98, 101, 102, 111]. Due to smartphones, some touch-based gestures have quickly become almost universally-applied and accepted, but most in-air gestures continue to fall under the "common knowledge" mentioned above. What seems not to be generally understood by the designers of gestural interaction is that gestures are not normally and intuitively applied on their own. They are usually either a compliment to speech (as in natural conversation) [120] or a learned action that serves a specific meaning under a specific circumstance (as in the sign language used for scuba diving [8]). The challenge, then, was twofold. First we would have to find gestures that already have well-known meanings; meanings that could be made to correspond to the items from the list of use-case-derived action words we were proposing as commands. Secondly, we would have to develop an easy-to-use means of capturing the gestures as computer input. We decided that since smartphones are becoming ubiquitous, an app would be an ideal delivery system, and we derived gestures based on universally-recognized actions that would fit in with the specific mental model of a magic wand; a cross-culturally recognized, hand-held device used for gestural commands. Technical aspects of the app and a summary of our pilot test are offered in Chapter 4.

6) Addressing Q5, we propose a system for intuitively interacting in an intuitive, multimodal, and "Calm" manner with a Smart Home: the C.A.S.A. T.E.V.A. app.

   With theoretical foundations in place, and preparatory studies complete, we address the issue of bringing all of these ideas together to answer the original question with real-time, in situ, empirical testing. We will conduct these trials using our C.A.S.A. T.E.V.A. app The zAPP App (see contribution 5) was modified and expanded for use as a portable, multimodal input and output device for smart home interaction. This included incorporating a personal major domo or butler [131], not only as a middleware [35, 65, 81, 126], but designed as a modified chatbot that would serve as a buffer or insulator from techno-stress [5, 79, 110, 88]. Specifically, the app is intended to give the impression of a single holistic concept of the home as a foundation of a new interaction paradigm. The new app, C.A.S.A. T.E.V.A. (Customizable Activation of Smart-home Appliances Through Enhanced Virtual Assistants) was the basis for real time, in situ testing that took place in June and July of 2013. The app and this testing are discussed in more detail in Chapter 5.

7) Addressing Q2, we propose the prototypical iteration of our "Measure of Calm".

Now that we have generated a practical model of Weiser's "Calm", we can address the long-standing issue of whether or not it is quantifiable. Weiser's idea of redesigning computer interaction to better suit the ways that humans naturally interact with the world has been an active topic of discussion for 20 years [152, 153, 151, 70, 22]. Despite that, there has been no attempt to develop a standard means of measuring that suitability. The complex nature of human performance requires an approach based in part on fields as diverse as human factors, the psychology of memory and attention, and neuroanatomy. Our prototypes are simple matrices that allow the simple classification of the attentional demands involved in any task; the C.A.L.Matrix (Classification of Attentional demands in a Layered Matrix). Two versions of this matrix are introduced in detail in Chapter 3. A related prototype, the S.H.I.E.L.D. (Simple Hazard Identification through the Evaluation of Layered Displays) is discussed under future work in Chapter 6.

1.5 Overview

The questions and answers introduced above are set in context in an overview of related work in Chapter 2. Here we will focus on reviewing the literature on selected aspects of Smart Environments (2.1), Human Computer Interaction (2.2) and Calm Technology (2.3). The holes uncovered in the related work become the footings for the theoretical foundations of this thesis. These foundations are presented in Chapter 3 as expansions and integrations of the international lectures and invited talks at which our theories were developed and shared for the first time. The development of *Calm Technology* for *Human-*

Computer Interaction in *Smart Environments* led us to propose the approach called *Anthropology-Based Computing*, which informed our *Brown-Hitz Model of Calm Interaction*, which in turn, pointed towards our prototypical *Measure of Calm*.

Concurrent to the formalization of those theories, we built prototypes and conducted a series of preparatory studies. Summaries of these published and as-yet-unpublished studies are summarized and presented in Chapter 4.

The ultimate experiment that brought together the culmination of our technological contributions in order to test our theoretical contributions is reported fully in Chapter 5, including qualitative and quantitative results, conclusions, and a critical discussion.

The dissertation will be concluded in Chapter 6, with a review of the contributions, critical reflections on our work over the past three years, and some proposals for future work.

References will follow.

CHAPTER 2

Related Work

As mentioned above, this dissertation is an attempt to bridge the rift between the technological aspects of Ubiquitous Computing and the theoretical or philosophical aspects of Calm Technology. Since it is unwise to build a bridge without a firm anchoring on both sides, we offer a review of the literature of some interrelated fields, specifically focusing on selected aspects of Smart Environments, Multimodal Interaction, and Calm Technology.

2.1 Smart Environments

From cockpits and nuclear power plants to the average 21st century home theatre or media center; from virtual reality-augmented surgical theatres to immersive-gameplay arcades; smart environments are no longer restricted to science fiction. In fact it can be postulated that, given the ubiquitous use of smartphones, their expanding toolset, and the almost universal nature of connectedness, we now carry our smart environments with us. Ubiquitous computing has turned not only our homes and workplaces, but even the most prosaic environment (a train car or a city park) into a node in a network of embedded systems. The average person may not even be aware of the degree to which they are connected. This is the situation predicted by Weiser in 1991. Since the dawn of the internet and the beginnings of incidental connectivity, our proclivity for connectivity and our demand for service have surpassed all predictions. One area in which this proclivity has been a driving force is the domestication of the technology behind smart environments. We will now review history and state of the art of this subset of smart environments, the Smart Home, before returning to the more general field to conclude our discussion.

2.1.1 The Smart Home

Research into smart homes has been going on for decades and detailed reviews of the literature have been conducted by Cook and Das [47], by Chan et al. [37], and more recently by De Silva, Morikawa and Petra [55]. The focus of these studies is often on Ambient Assistive Living (AAL) for the elderly [56] or for people with special needs [17, 112], but the entry threshold for AAL is dropping with the advent of innovative design and technology integration [111]. This is changing the nature of smart environments, especially as technological advances allow display and control to change from single-user to multi-user [67].

Leonardo gave us what may be the first documented transcription of technological innovations into normal living space in his folios numbered 16r and 37v, as seen in the collection of codices of the *Institut de France* [114]. It wasn't until the early 20th Century that the modern concept of a home technology entered the public consciousness, largely in the form of comedy in which incredibly-complex, automated, Rube Goldberg-style machines were proposed as a means to "simplify" day-to-day tasks, such as the feeding machine in Charlie Chaplin's "Modern Times". The advent of practical computer technology in the middle of that century, and the creation of the first commercial microchips led to the beginnings of home automation. Simple devices such as remote controls for televisions and garage door openers were quickly accepted internationally [149]. More complex devices generated publicity but not sales. One example is the "Honeywell Kitchen Computer" (H316 pedestal model) offered in the Neiman Marcus consumer catalogue in the United States in 1969 [10]. This machine was advertised as being able to help housewives plan their menus and budgets, but it was roughly three times as expensive as a house and required that the user take a two-week course in order to learn to use the toggle-switch input panel and to read the flashing binary light output.

Both price and demands on the user would have to be lowered before computerized assistive devices could become realistically viable in the home. It is interesting to note that, according to Atkinson, Gordon Bell (Vice President of Engineering at Digital Equipment Corporation (DEC)) wrote a memo describing possible improvements to the "Honeywell Kitchen Computer" that was the inspiration for the DEC to enter the field of domotic computing. In fact, in the memo, Bell wrote that, with an improved interface, a home computer could be directed not at the kitchen but for use with entertainment, games and studying [16]. By the end of the century, Smart Home systems were being developed and tested in academic and corporate laboratories around the world [37].

One of the major drivers in the quest to build Smart Home technology is the profit motive of private developers who show little concern for how the technology will be used or whether it is compatible with other technology. In 2002, Zayas-Cabán proposed a methodology for conducting home assessments in order to implement specialized technological systems that suite the house and the inhabitants [159]. As discussed a decade earlier in the Report of the European Foundation for the Improvement of Living and Working Conditions, the manufacturers of technological devices were putting the cart before the horse: rather than assess the environment and develop technology in response to needs, they were waiting until after technology was developed and deployed before worrying about how suitable it might be [128]. This has led to a vast treasure trove of commercial systems and components and to a dearth of commercial attempts to work with other developers. The void is being filled now, at least on a theoretical level, and many academic and commercial research teams are turning their attention to finding not just models, but actual working systems for the unification of commercially and technologically diverse distributed Smart Home interfaces.

Many of the Smart Home systems developed to date have two unfortunate elements in common with the kitchen computer discussed above: they are very expensive, and they make high demands on the user. Both of these elements were addressed in the Casa Vecchia project [112, 113].

2.1.2 Casa Vecchia: Making an "Old House" Smart

Leitner and Fercher developed Casa Vecchia, a Smart Home project that is outstanding, not only for addressing issues of cost and cognitive load; and not only for setting out to evaluate the viability of deploying Smart Home systems in their community; but mainly because it deals with the oft-ignored guideline suggested by Venkatesh:

> *"Don't assume that what the technology can do in the household is the same as what the household wants to do with the technology" [150].*

After a decade of conducting anthropological-style field studies and large-scale longitudinal and cross-sectional surveys in order to determine the technological and social elements affecting technological diffusion in the home, Venkatesh developed an underlying theoretical structure that included a conceptual model of the *cyberhousehold* [149]. His theory of household-technology interaction is generated from a modified structural-functionalist approach that has sound footings in ethnography but is largely ignored by the technological community.

Leitner and Fercher, like Venkatesh, approach the Smart Home from the points of view of both utilitarian material culture (focusing on tools and tool use) and a socio-psychological approach in which the social dynamics of the household must be paramount. By combining these perspectives, Venkatesh was able to model the use of technology in relation to household structure, and so propose a dynamic and adapting system that would change according to the needs or wants of the occupants of the home. This dynamic, human-focused application of technology is exactly what Leitner and Fercher set out to apply and study, in the hopes of "...finding as many missing pieces of the jigsaw of UX in the context of AAL as possible."

They have done this by combining HCI approaches to the human side of the equation with innovative applications of off-the-shelf technology while standing on the shoulders of those who have gone before them. To use their words:

> *"The focus of the project is to deploy a customized system into more or less arbitrary homes based on the achievements gained by researchers all over the world" [112].*

The homes in question are the 21 real houses of real people in the region surrounding Klagenfurt, Austria. The senior citizens living in these homes have agreed to a slow introduction of domotic technology. They, their families and their primary caregivers are all involved in an ongoing process of feedback and response as the researchers and the participants co-develop the customized systems that best suit the household. The results of longitudinal surveys will soon be available. In the meantime, though only preliminary results have been reported so far, this truly human-focused means of developing a smart environment was instrumental in fostering the theoretical and practical work reported in this dissertation.

2.1.3 Computer-Centered Computing

In January, 2011, the US National Science Foundation gathered a group of 72 international researchers in Seattle to discuss the multidisciplinary problems involved in the future of networking smart tools. The discussion is summarized by Cook and Das [46]. The workshop and resultant paper focus on scalability as the key issue for the future. They broke down their concerns into eight subfields, only one of which directly focused on the human factors in HCI.

This focus on the machines rather than the people who should use them is a weakness in current and past trends in pervasive computing, despite the relatively long history of applying the perspective of cultural anthropology to the adoption of cyber technology [64]. It is a shame to think that this will continue into the future, but consider the opening sentence of the summary paper mentioned above:

> *"The remarkable recent growth in computing power, sensors and embedded devices, smart phones, wireless communications and networking combined with the power of data mining techniques and emerging support for cloud computing and social networks has enabled researchers and practitioners to create a wide variety of pervasive computing systems that reason intelligently, act autonomously, and respond to the needs of the users in a context- and situation-aware manner." [159]*

The idea that intelligent agents should be making hidden decisions on behalf of humans is totally against the idea of UC mitigated with CT as envisaged by Weiser. Despite that, and despite Weiser's direct warning, as cited above, researchers have continued in their attempts to generate intelligent agents intended to make decisions so that humans don't have to. Consider Diane J. Cook's monograph in the March 2012 issue of Science [45], wherein she expresses an indiscriminate overlap between pervasive computing, ubiquitous computing and ambient intelligence, positing a home or work environment that is entirely under the control of intelligent agents. This is a far cry from the "gentle enhancement" of the natural environment with "self-effacing" interfaces that would "leave you feeling as though you did it yourself" posited by Weiser [157]. In fact, the idea of smart environments based around intelligent agents seems to be the inverse of Weiser's idea of "[m]achines that fit the human environment instead of forcing humans to enter theirs" [155].

If it is accepted that using computers causes stress when the user feels that they are not in control, as per Riedl et al. [136], then it is a natural extension to assume that such stress would be an even greater threat in an immersive, computer-centered environment such as a smart home. Interviews and focus group sessions have shown that users prefer a centralized remote control to enable immediate interaction with a number of devices installed in a household [110]. While the concept of a control panel proved popular, as an interface, it is an artifact from what Weiser called the *Mainframe Era* of computing [153].

While some researchers, such as Chan et al. [37], foresee the coming of either wearable or implantable systems to complement domotic control with the provision of biomedical monitoring sensors, it will be some time before these features can become ubiquitous. They go on to stress that since smart homes promise to improve comfort, leisure and safety the interaction should be as natural as possible. If their proposed method of improvement is still developing technologically, our proposed method is built upon applying currently available technology in a novel manner.

2.2 Human-Computer Interaction

Early human-computer interaction was a multi-stage process, requiring that several specialists work on a single project. Those requiring computer assistance would consult these specialists, whose skill was the ability to communicate with the machine. Since the machine, in those days, was essentially a series of on/off switches, all of the input mechanisms provided serial information; hundreds, thousands, even millions of noughts and ones.

The first area of specialization was the translation of the question into problems that could be presented to the computer. A question would have to be expressed as a series of logic problems, the sort that could be answered "yes" or "no". The series of questions framed by the logicians had then to be translated into "machine language" by a group of translators, and then passed on to experts who created tapes or punch-cards. These were then passed on to the technicians who actually worked with the machine. According to one account:

> *"This seems vastly more complex than the computer systems that we use now, but the only real difference is that most of the specialized steps in the process of human-computer interaction are now performed by the computer, rather than by the human."* [20]

To paraphrase Myers [122]; the change that allowed HCI to move from a field for experts to a field for common use was the realization that, through the addition of processing power, the machine could assume most of the expert roles. This was a great breakthrough in the proliferation of computers into day-to-day life. Unfortunately, as shown from the continued use of obsolete 400 codes for flagging errors (like the common "Error 404"), it gave the world a working model of computer-centered interaction that we have still not overcome. To evolve past computer-centered computing, one necessary step will be to stop our *Cross-Generational Habit* of designing interaction in accordance with obsolete technological standards like typewriters or TV screens [23]. In its place we should establish new human-centered standards based on a better understanding of the natural workings of our brains and on simple, observable facts. One example that is pertinent to HCI is the observable fact that human communication naturally involves complimenting words with gestures.

2.2.1 A Gesture of Goodwill

> *"In gestures we are able to see the imagistic form of the speaker's sentences. This imagistic form is not usually meant for public view, and the speaker him- or herself may be unaware of it…"* [120]

As we have already said, all natural human interaction is multimodal; we constrain ourselves to a single modality only when required. When in a diving environment, scuba gear enables us to function without having to learn to breathe underwater, but formal communication is reduced to a single modality and becomes dependent on the use of strictly-defined and well-practiced gestures. When in a digital environment, the GUI interface enables us to function without having to learn machine language, but formal communication is reduced to a single modality and becomes dependent on the use of strictly defined and severely truncated words which have been removed from their usual ontological, cultural and environmental context.

Hurtienne et al. produced what they claim is “the first study looking into primary metaphors for gesture interaction in inclusive design” [91]. Their paper proposes the construction of physical gestures, based on the aggregate of twelve of what they called primary metaphors from other published studies. This list is quoted directly from theirs:

1) Important is central, unimportant is peripheral.
2) The future is in front, the past is behind.
3) Progress is forward movement, undoing progress is backward movement.
4) Similar is near, different is far.
5) Familiar is near, unfamiliar is far.
6) Considered is near, not considered is far.
7) Good is near, bad is far.
8) Good is up, bad is down.
9) More is up, less is down.
10) Happy is up, sad is down.
11) Virtue is up, depravity is down.
12) Power is up, powerless is down.

In each case, the authors quote spoken phrases that were used to support the metaphor in the original paper. At best these examples are facile, as in the pair “I feel close to him. He distances himself" used to illustrate number 5, which is easily countered with the common phrase "stranger in a strange land". At worst, the chosen phrase does not mean what the authors seem to think. Consider the phrases used to support metaphor number 7: "Here is something interesting. There comes the difficulty." "Here" and there" are interchangeable in the first phrase, and the second phrase would only be correct English if "there" were replaced by "here". This linguistic confusion is unfortunate, but it does not lessen the problem of trying to base a universal gesture on a non-representational subset of world languages. It is possible to generate lexicons of language- or culture-based gestures, as we have seen in the work discussed above. It is also possible that there will be some overlap between these gestures and any new lexicon of truly universal ones. Such an overlap, however, may be much smaller than one would initially anticipate.

If one were to pursue the idea of universal gestures, which is not the intent of this dissertation, it might be better to turn away from simplified contrasting pairs like "up" and "down". Consider that up and down can both mean, "at hand", "within easy reach", "within difficult reach", "out of reach" and "far beyond reach".

High or low, "out of reach" has a meaning that is universal. It is different from "within reach", and both are different from "in-hand" and "unreachable", but none of them are opposites. Neither are "hot and cold". "Too hot" and "too cold" can both be mapped as cognitive vectors away from a concept of comfort. Maybe we can agree that they are both going through the realm of "discomfort", towards a concept of "environmentally fatal", but these would clearly all be best conceptualized as concentric spheres rather than 1-dimensional lines.

Leaving aside these logical flaws, Hurtienne et al. claim that their metaphors have not been influenced by technology. This claim is refuted in their examples and through simple inductive reasoning. To wit: the experimenters and participants have all been influenced by underlying mental models in their most basic technological tools, such as the Cartesian increase in value when a switch is pushed either to the right or upward, and the decrease when the same switch moves in the opposite direction. Since these underlying concepts are applied globally, it would seem obvious to save time and trouble by simply using them as the basis of gestural interaction. Our attempt to do so, is described in 5.6.2.

The technology that supports gesture detection has greatly dropped in price and increased in popularity with the advent of gesture-based video game interfaces. This started with actively-broadcasting sensors in handheld gameplay controllers and active motion detectors in stationary consoles. Further improvements in the availability of gestural interaction have come from the development of smartphone applications that take advantage of the increasing presence and improved performance of accelerometers, magnetometers and gyroscopes. Despite the usability afforded by the increasing ubiquity of smartphones, empty-handed interaction is still a goal. Cohn et al. proposed a means for using the electromagnetic field generated by normal in-home wiring to detect the location, orientation, and hand and arm movements of participants in their own homes [43]. Refinements to their method, *Humantenna*, were presented the following year, with results suggesting multi-finger gesture recognition [44]. In both cases, though, the participants had to wear a backpack full of equipment. Our own proposal for smartphone-based and empty-handed gestural interaction will be presented in Chapters 3, 4 and 5.

We do not want to create a gestural recognition system based on the false paradigm of single modality interaction. Instead, our gestures will amplify spoken word interaction, observing the same phonological synchronicity rules that have been observed when gestures accompany speech in normal interaction [120]. Rather than succumb to the common behavior of designating a black box to encapsulate technological issues that have not yet been resolved but do seem to be imminently soluble, we have turned to an old triple-redundancy protocol that was used a generation ago on satellite control systems [95]. Our attempt, the S.N.A.R.K., is mentioned again below and discussed in detail in Chapter 4.

2.2.2 Speech and Sound

The control of networked and embedded systems through the use of automatic speech recognition has long been a feature of science fiction and fantasy interfaces, but the idea of implementing the technology in the real world predates the modern computer era, as reflected in the first volume of Natural Communication with Computers, from Woods et al. in 1974 [158]. Even with the development of superior technology, attempts to translate the idea into real life have met only modest success, as is reflected in the title of a 2011 conference presentation by Oulasvirta et al., “Communication failures in the speech-based control of smart home systems” [127].

In 2008, Fleury et al. presented a study of speech and sound detection and classification (n=13) that took place in the Health Smart Home in Grenoble, France [73]. The study is based on the use of 8 ceiling-mounted, omnidirectional microphones that are always turned on. Participants provided data in 3 ways. First they were asked to “make a little scenario” that involved closing a door, making a noise with a cup and spoon, dropping a box on the floor, and screaming the common French exclamation of pain “Aie”. This was repeated two more times. Next, each participant had to read 10 “normal” sentences and 20 “distress sentences”. Finally, the participant read a conversation into a telephone. Random selections from these separate noises and phrases were then chosen for testing the system. The authors report that the sound recognition results conform to results from laboratory conditions. Speech recognition results were “too low”, with 52.38% of the noises made with cup and spoon, and 21.74% of the screams of pain and 62.92% of the distress speech recognized as normal speech. The authors propose that one of the difficulties in speech recognition is that each participant pronounced each phrase differently each time they repeated it. Further weaknesses identified by the authors include background noise, the freedom of the participants to work at their own pace, and the uncontrolled orientation of the speaker vis-a-vis the mounted microphones. Hurtienne et al. sum up these weaknesses very well: “Thus our conditions are the worst possible, far from the laboratory conditions (no noise and the microphone just behind the subject)” [91]. These weaknesses were used as guidelines for developing our own testing as reported in Chapter 5.

In 2009, Hamill et al. also used ceiling-mounted microphones in their proposal of an automated speech recognition interface for use in emergencies [82]. They compared results with a single microphone to an array, in both noisy and quiet conditions (n=9), and they tested a yes/no response dialog with four participants. Their array performed with 49.9% accuracy, and the single microphone with 29%. Recognition of the words “yes” and “no” was 93% accurate over their 3 scenarios, even though overall word recognition had an error rate of 21%. The researchers suggest that the “reason for this was because the system confirmed the user’s selection before taking an action.” The authors also report that background noises interfered with the performance of their microphone(s) and that speech recognition was greatly improved by limiting the user’s speech to two words. Again, we have been inspired in the design of our experimental protocol, to try and face the specific problems described herein. We were also influenced by another aspect of this study. In the discussion of future work, Hamill et al. mention that, in order to improve the robustness of their automated dialog system, they are developing a speech corpus recorded by an older adult speaking Canadian English. We went on to do the same.

In 2010, Chandak and Dharaskar reported an attempt to implement speech-based controls for a context-sensitive, content-specific Smart Home architecture based on natural language processing [38]. The key to their system was the ability of the user to customize the specific language or languages to be used for input. The paper itself seems incomplete, presenting none of the results promised in the introduction. In fact, no evidence is provided to indicate that the system was implemented at all. That said, the premise of customizing input language on a dynamic, user-by-user basis informed our theoretical development and the implementation discussed in Chapter 5.

Two teams in France, GETALP in Grenoble and AFIRM in La Tronche, undertook an attempt to design a real-time smart home distress-detection system based on audio technology [148]. They started by testing speech-based detection of distress using a scenario based on a prepared corpus of phrases (n=10) and reported an overall error rate of 15.6%. For their second experiment, four participants uttered prepared "distress sentences" while a radio news program played in the background. Distress went undetected 27% of the time. In 2010, the same research teams attempted to apply their more advanced sound and speech analysis system (AuditHIS) to the recognition of Activities of Daily Living (ADL) [74]. They attempted to validate their stress-related keyword detection and their algorithm for suppressing background noise while using AuditHIS (and installed sensors) to identify 7 ADLs. They again conducted their experiments in the Habitat Intelligent pour la Santé (HIS) Smart Home, in Grenoble, a site that they describe as "an hostile environment for information acquisition similar to the one that can be encountered in [a] real home". Specifically, they note that uncontrolled noises outside and around the flat reduced their average signal to noise ratio to 12dB, from the 27dB measured in their laboratory setting. These normal, uncontrolled, noisy conditions inspired us to address the same issue in both a preparatory study and our final experiment.

As part of France's new Sweet-Home project, Lecouteux, Vacher and Portet compared 7 sources of Automatic Speech Recognition (ASR) [109]. Twenty-one participants recorded pre-determined phrases. Each acoustic model was trained on "about 80 hours of annotated speech". In the end, they report that the array of seven microphones improved ASR accuracy and that Beamforming (as in their previous experiments) dropped the Word Error Rate (WER) from their baseline of 18.3% to 16.8%. They found that a Driven Decoding Algorithm (DDA) had only a 11.4% WER and provided slightly better results than the SNR-based ROVER system. Since the computational cost of the DDA is significantly less, and since the DDA would allow for the inclusion of *a priori* knowledge parameters which would significantly improve the results.

In 2011, Gordon, Passoneau and Epstein presented FORRSooth, a multi-threaded semi-synchronous architecture for spoken dialog systems as an improvement over CheckItOut, their previous pipeline-style architecture [79]. They reported on a pilot study suggesting that helping agents are helpful even when their speech recognition is not perfectly implemented. They addressed the important problem that most ASRs do not allow people to speak naturally during interaction. They suggest that a spoken dialog system (SDS) "should robustly accommodate noisy ASR, and should degrade gracefully as recognition errors increase." This would allow "more nuanced grounding behavior from an SDS" and "help a system understand its user better." These ideas supported our intent to create a dialog system that would help both the software and the user to understand each other better. This nuanced system is described in Chapter 5.

A different kind of sound-based output signal is reported by Bakker et al. [14] They propose CawClock, an interactive system designed to allow a schoolteacher to set peripheral audio and visual cues by placing tokens depicting distinct animals and colors on the face of an analog clock. These placements cause sections of the clock's face to match the color of that particular token. So long as the minute hand is within the colored area, a background noise that corresponds to the animal is generated by the clock. The volume does not change but the number of animals making the sound increases as time passes, providing subtle cues that time is passing and that the end of the particular timeframe is approaching. Two prototypes were developed. An analog model was provided to the teacher in one classroom for 2 weeks. A mouse-enabled digital version was provided to another. As a reflection of the exploratory nature of the study, the teachers were taught how to use the device but asked to find their own uses for it. Two researchers then attended a 30-45 minute classroom session during the second week, taking notes and recording the class on video. The teachers were interviewed singly and together at the end of the second week. Both teachers agreed that the sounds gave themselves and the children signs of passing time during periods of assigned work in the classroom. Both teachers also agreed that they had not noticed the increase in the number of animals over time. Interestingly, and in line with the fundamental understanding of peripheral perception, the ending of a marked period was reportedly more distinct when the background noise changed from one animal to another rather than when it simply ended.

Two unsolved questions in the realm of sound and speech detection have been whether or not to have constant sound detection and whether or not to have a live processor. This would mean a constant drain of both electrical power and processing power. Problems regarding processing power and the extension of battery life are easier to deal with quantitatively. The problems of accurately distinguishing sounds and recognizing speech are generally labeled "not insignificant" and replaced with a black box in flow charts and designs. As with our approach to resolving black box issues in gestural interaction, we have used an old triple-redundancy protocol as the basis of our S.N.A.R.K., a means of facilitating the accurate detection of user intent as communicated through natural means. This is discussed in detail in Chapter 4.1.

2.3 Calm Technology: "...as refreshing as taking a walk in the woods"

In 1991, Professor Mark Weiser wrote that "Machines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods" [155]. Current issues with interoperability, product design and human factors prevent Smart Home users from being able to see the forest for the trees, but this does not have to be the case. As mentioned earlier, Professor Weiser coined the term "Pervasive Computing" (PC) and theorized that it would eventually lead to an era of "Calm Technology" (CT) wherein computers would be embedded not only in our devices, but in our lives as well. As he put it:

> *"...the most profound technologies are those that disappear. They weave themselves into the pattern of everyday life until they are indistinguishable from it."* [*155*]

Calm Technology is based on the way that humans process information: the process of plucking things from the periphery and deciding how to prioritize them is a comforting activity that makes us feel at home and in control. Weiser and Brown provided examples of calm output from a network, but did not provide examples of calm input. That said, it seems logical that calm input would be based on the natural means of human to human communication. This requires a greater depth of multimodality than has previously been common.

2.3.1 Understanding Calm

Calm Technology describes any tool that can be used with uninterrupted focus on a central task while new outside information is easily perceived and processed peripherally [152]. This dynamic allows the user to decide whether to divert their attention and change their focus at any time. This is the natural means by which all primates interact with their environment [23], and it is a fundamental part of the iterative cycles of perception, evaluation, and reaction that have shaped our evolution and that continue to shape our understanding of, and interaction with, the world around us [50]. Furthermore, Calm Technology allows one to focus on their intended action rather than on the tool they are using [153].

The concept of Calm Interaction is often misunderstood as a means of calming the user, as at The Stanford Calming Technology Lab, and in Rogers' call to abandon the concept in favour of "engaging rather than calming people" [138]. As mentioned earlier, "Calm" has also been conflated with the automation of decision processes, as exemplified by Makonin et al. [118], Olaru et al. [126], and Stavropoulos et al. [143], despite the obvious fact that automated decisions remove control from the user rather than simplifying it.

The problem that "Calm" is intended to address is the problem of enabling people to deal with large amounts of information without becoming either overwhelmed by stress or oblivious to the world around them.

2.3.2 Is "Calm" Necessary?

Csikszentmihalyi's concept of flow [123] has been used to describe situations in which the demands of task performance so coincide with the abilities of the performer as to enhance performance rather than limiting it. This total immersion might sound like calm, but it should not be confused with "Calm" in the sense that Wesier described it. Being in "flow" reduces one's ability to perceive or interact with the rest of the world [51].

Santangelo, Fagioli and Macaluso have described the natural processes by which the brain deals with large amounts of multimodal information [140]. Our ability to deal with interruptions has been discussed extensively for most of the last century, with informative analysis of the processes and the stresses involved provided by Zeigarnik, in 1927 [160] and by Gillie and Broadbent in 1987 [78].

According to Brod, Techno-stress comes from feeling overwhelmed by information and options which are not fully understood [18]. Riedl et al. confirmed this by measuring the presence of stress hormones in saliva, and proposed that this stress can be mitigated by developing a feeling of being in control [136].

It was Weiser and Brown who suggested that the natural means by which humans sort and prioritize information in their natural environment could be the remedy for the stressful environment created by UC [153]. They postulated that following natural processes of prioritization would allow humans to feel more at ease and even to distinguish between natural situations that should or should not be classed as legitimately stressful. It seems logical to infer that deliberately-designed interactions could be created in a manner that is free of unnecessary or unproductive stress

We have proposed, in an earlier work, practical examples of "Calm" in modern computer interaction; input and output systems based on human information processing rather than machine processing [27, 28]. While this has been discussed theoretically for some time, by the likes of Polanyi [132], Popper [133], Cadiz et al. [31], and van Dantzich et al. [149], it has led to very few practical examples. Weiser and Brown suggested examples from the analog world, like semi-opaque inner-office windows [153]. Virolainen et al., suggested HCI devices [151] and Bakker et al., created peripheral interaction devices to help students track the passage of time [13] and to help teachers photo-document student activities in the classroom.

Bakker et al. also proposed qualitative measures of "Calm" based on post-hoc user surveys. Other qualitative metrics were demonstrated by Peterson [129] and Gaudron and Vignoli [76], but to date, no quantitative metric has been proposed.

In the next chapter we will attempt to resolve this issue by introducing the Brown-Hitz model of "Calm" HCI and a prototypical means of measuring whether or not a tool or task is "Calm".

CHAPTER 3

Theoretical Foundations

Now that some holes have been found in the related work, it is time to see if they can be used as footings from which we might build the foundation of a new realm of HCI. What follows is a discussion of each of the three theoretical innovations we propose to contribute to this foundation.

3.1 Anthropology-Based Computing

The chapter is based on the homonymous keynote lecture delivered at the 12th International Work-Conference on Artificial Neural Networks (IWANN 2013) in Puerto de la Cruz, Tenerife, Spain, on June 12-14, 2013. A monograph based on the lecture was published as:

Brown, John N.A., 2013, Introducing Anthropology-Based Computing: It's as Easy as ABC, in Advances in Computational Intelligence: 12th International Work-Conference on Artificial Neural Networks, IWANN 2013, Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013, Proceedings, Part I Lecture Notes in Computer Science 7902, Ignacio Rojas, Gonzalo Joya, Joan Gabestany (eds.) Springer Berlin Heidelberg, pp 1-16.

The ideas were further developed for an invited talk at the Delft University of Technology on July 11th, 2013.

Anthropology-Based Computing is the application of the fundamentals of Anthropology in order to remake traditional Human-Computer Interaction into a science that is truly based on humans, instead of the motley series of brilliant innovations, glorified mistakes, and obscure *Cross-Generational Habits* that comprise the computer-centered HCI that we practice today.

Currently, a single human uses dozens of computers, or maybe hundreds or even thousands, if one includes the machines that are used by many individuals, such as: the servers run by Google or by Wikipedia; the network that helps you find a flight, or; the one that supports the television weather forecaster in their nightly performance. This is, of course, in addition to the computers or computerized systems shared at work, at home and during the transition between the two. One should also include in the list any and all personal systems being used either deliberately or without any conscious awareness at all. After all, being consciously unaware of a pervasive technology is a sign of being well-adjusted. Less than a century ago, having electricity in the walls was a fantastic idea that could not be ignored. Fifty years ago, the same was true of using a computer. All that changes from generation to generation is the technology we learn to ignore.

To explain this perspective and establish a historical foundation for ABC, let us consider what human-centered interaction really looks like in the natural world. This has been developing for a long time.

3.1.1 The Proto-Prosimian and the Workstation, Part 1

About 45 million years ago long-tailed and long-fingered prosimians could look at a single object with both eyes at once, giving a sense of depth perception, and establishing that some things are near and others are further away. In the midst of a rich sensory environment, we had to learn to focus on single tasks without losing our perception of the cascade of environmental input. How would we do that?

About 30 million years later (about 14 or 15 million years ago), we no longer had a tail and the fingers on our feet had become a little shorter than those on our hands. Our faces were flatter, and we had started to spend more time out of the trees and in the tall grasses of what is now Southern Europe. Our cerebral neocortex was now more developed than that of the previously-discussed prosimian, which probably allowed us to better compare possible outcomes of our actions. The question remains, how would we focus on one task while maintaining peripheral awareness?

Approximately six or seven million years later, our ancestors follow the receding warmth and migrate into what will later become Africa. We are bigger now, with a much bigger head and a much bigger brain, but facing the same small problem. We know that this ancestor had delicate, precise, and very powerful fingers, and was as related to modern humans as to chimpanzees and bonobos. But what do we know about how they interacted with their environment?

The truth is that we don't *know* anything about what happened then. We can only apply inductive reasoning and try to generalize a theory based on the fossil record and the behavior of similar animals today.

We may not know what they did, but we do know what they didn't do – what they could not possibly have done. Not one of our remote ancestors, from this series of examples or from any other, could possibly have approached the problem facing them by thinking like a computer.

If we are going to consider the issue of Human-Computer Interaction in any kind of a meaningful way, then we must remember that humans come from stock that spent millions of years not thinking like computers. This is the basic fact that Weiser was trying to communicate when he stressed that the ubiquitous presence of computers in our lives would make it absolutely necessary to change the way they work. Weiser wrote:

> *"Calmness is a new challenge that UC brings to computing. When computers are used behind closed doors by experts, calmness is relevant to only a few. Computers for personal use have focused on the excitement of interaction. But when computers are all around, so that we want to compute while doing something else and have more time to be more fully human, we must radically rethink the goals, context and technology of the computer and all the other technology crowding into our lives. Calmness is a fundamental challenge for all technological design of the next fifty years." [153]*

The goals of ABC are to explain why Weiser was right to say that HCI must be changed when computers become ubiquitous, and to show how that might be done

If alarms and alerts annoy you when they come from a smartphone at a nearby workstation, park bench or seat in a movie theatre, how much more would they annoy you if you were living inside the phone? The smart environment of the near future will surround us with computerized recommender systems, ambient information systems, and distributed interfaces and displays. No one expected that the widespread dissemination of electronic mail systems would lead to incessant interruption or that text messaging functions on portable phones would mean that teenagers would be in constant low-fidelity communication throughout their waking day. How will incessant communication expand when every wall, window and door of our homes is automated? How will our "time to be more fully human" diminish when every device in and around our lives is always on the verge of demanding that we stop everything and reply to the 21st century equivalent of "Error 404"?

3.1.2 Inhuman-Computer Interaction

Many research teams around the world are working on the technological side of that very issue – finding ways to make networked and embedded systems with which humans might surround themselves; systems that can anticipate human requirements and enrich our lives. But they are doing so with the same perverse idea of Human-Computer Interaction that dominates other parts of the industry. The flaw in their reasoning is obvious, but most of us are simply choosing not to consider it.

Modifying a device so that it becomes less harmful to the user is a vital step in the early evolution of any tool. This is one of the reasons that our ancestors added handles to axes. It was a technological improvement, in that it increased the length of the lever arm and made possible a series of adaptations that led to further tool specialization, but another major part of the improvement is that a handle made it less likely for the tool user to hurt herself. There has always been an accepted trade-off between the danger and the value of using a tool. If a tool is too dangerous, but must be used, society tends to restrict its availability. Eventually the tool is improved, given up, or moved entirely into the realm of specialist use. These matters take time, but the amount of time is dependent upon societal perception of the dangers involved.

3.1.3 Computer-Centered Computing

Consider the surface modifications of the computer mouse, the desk chair, or the computer keyboard in response to societal perception. These small "ergonomic" adjustments do not address the fact that none of these tools are really designed based on human needs and abilities. In fact, these tools are all examples of machines that force humans to enter their world. Making the keyboard ubiquitous changed the rare medical condition that used to be called "Secretary's Disease" into the extremely common Carpal Tunnel Syndrome. A tremendous amount of time and money is spent in the attempt to make sitting at a desk less strenuous for office workers, and this industry has followed the computer from the office to the home, but the truth is that humans evolved as walking creatures. It puts our bodies under strain to sit upright for an entire working day. Moving a computer mouse on a desk requires constant interruption of anything else one might be doing with that hand, and puts the wrist and forearm under stress that is very similar to that caused by the keyboard.

Muscular contraction or resting weight puts pressure on the nervous, circulatory, and musculo-skeletal systems of the limb, and precise and repetitive movements while under pressure causes repetitive strain. It can cause discomfort, numbness or outright pain to people who are using the device according to the manufacturer's instructions. So why do we keep using tools that are not good for us? What is the scientific explanation of *Cross-Generational Habit*?

3.1.4 Big Skulls, Brain Development, Culture, and Conformity

Genetically modern humans (GMH) have unusually big and complex upper brains, and so require unusually big skulls. There are a number of interesting theories about the way in which our brains were pressured to grow. Some say that our increased brain size is due to the proliferation of broad-leafed trees and other vegetation during the late Mesozoic Era, which increased the available amount of oxygen in the atmosphere. Others theorize that extra thickness in the cerebral cortex was an evolutionary response to our need for insulation from heatstroke when we moved into a diurnal life on the African savannah.

A skull big enough to house a fully-developed GMH brain cannot fit through GMH pelvic openings, so we are born with a brain that grows and develops over years. GMH are born with a brain only 25% of adult size. Our closest cousins in the modern world, bonobos and chimpanzees, are born with brains roughly 35% of adult size. By comparison, the brain of a capuchin or a rhesus monkey is already close to half adult size at birth and their skulls have stopped growing. A GMH brain takes about eight years to reach full size and then spends about another eight years (or more) maturing. The result of this is that our young need to be protected for at least eight years while they develop their mental abilities. This period of learning how to think while learning what to think could be responsible to some degree for the richness of human culture and language. If our brains learn to think while exposed to a particular home environment, wouldn't we naturally develop an understanding of the world, of how things should and must be, based on that source? If the familiar gives us comfort, couldn't some of that comfort come from the fact that a familiar environment makes it easier for the unusual (and possibly dangerous) to stand out?' Thus, pattern recognition and, more fundamentally, the ability to detect sudden changes in regular patterns becomes an important survival strategy. Let us look even further into the past to imagine how this might have been applied.

3.1.5 The Proto-Prosimian at the Workstation, part 2

Looking twenty-million years further back than before, we can see a cat-sized proto-prosimian sitting in a tree. The pointy little snout is sniffing at food. Since this is before the era in which flowering plants and succulents will spread widely around the world, it is unlikely that our hero is holding a piece of fruit. Let's assume, in its place, a nice juicy insect. Without binocular vision, our hero's head must be cocked so that one eye can focus on the lovely snack. The long tail twitches, and our hero thinks about this snack without losing track of everything else going on in the environment.

The proto-prosimian is dealing with the situation in a very human way, processing information in a manner that is probably very similar to the way that his descendants in our earlier examples would have done it. The little proto-prosimian, closer in shape to a modern mouse than to a modern man, is using the natural abilities of brain and body to deal with the problem. The cerebral neocortex, the part of the brain that lays like a wrinkled blanket overtop of all the rest, would likely have been very small in the skull of our proto-prosimian protagonist. It would be small, that is, in comparison to modern humans, but vastly bigger in comparison to earlier creatures, if it had existed in their skulls at all. Using two separate parts of the brain simultaneously, the creature is processing input in two different ways. Let's try to imagine it in more detail.

You are a proto-prosimian sitting in the crook of a branch about sixty-five million years ago. Your hands have fingers that curve only inwards, and your arms reach only about as far backwards as your peripheral vision can see. At rest, your arms fall to your sides with bent elbows and your hands overlap in front of your chest. I believe that this is your region of focus. Your precise little fingers overlap here and hold things where you can best smell and taste them. It is also the area in which you can most easily focus your eyes on near objects. Of course, this is where you want to hold your food, so that you can really focus on it. At the same time, though, you are aware of your surroundings. Your ears point outwards, shells cupping and adding directional information. Your hair detects wind movement and you sweep your tail back and forth, to add to your chances of early detection of shifting air currents. Our senses work together to form perceptual units out of these data, grouping them according to characteristics like spatial continuity, chronological coincidence and symmetry. Every now and then you pause, looking away or tipping your head to one side. These interruptions of your routine happen when you have detected something on the periphery of your attention, something that doesn't seem to fit an anticipated pattern and so might become important; something that you might need to consider more deeply. You weigh the importance of interrupting your meal and you either return to eating or drop your hands a little and stop chewing so that you can divert more of your cognitive resources to processing the information. This is normal behavior.

This is how we feel comfortable, surrounded by large slow streams of perceptual data, most of which we feel safe to ignore. Though we do not focus our attention on all of these currents of information, we feel their comforting presence around us and we believe that we can access them at any time, shifting our attention, and reassuring ourselves that all is well.

We have processed information this way for the last sixty-five million years. There have, of course, been situations where something demands immediate attention. Such situations, if they derive naturally, and if we have the opportunity to influence our own chances of survival, must trigger responses as quickly as possible. It is a survival characteristic to be able to respond quickly to stimuli that demand attention, just in case it turns out to be a matter of life or death. Similarly, it is a corresponding survival mechanism to avoid false alarms that might needlessly reduce our resources for dealing with real threats or might even desensitize us to stimuli that will be important later.

3.1.6 As We May Think

In 1945, the Atlantic Monthly published a monograph entitled “As We May Think” [29]. In this article, Vannevar Bush described the Memex, a special and entirely theoretical desk of the future. This desk would be able to receive, store, display, edit and delete information from around the world. His description may well be the reason that the developers of the first commercial GUIs used the analogy of the desktop. That analogy is central to our most-used computerized devices – the laptop and the smartphone. The thing is, when Bush proposed this possible future technology he wasn’t doing so just for the sake of talking about technology. Bush was saying that this access to huge amounts of data would affect our cognitive processing, changing us “as we may think”. We have not followed his example, but have slipped into the pattern of developing new technology without worrying at all about how the use of it might affect human thought.

Until very recently, writing to someone usually meant waiting days, weeks or even months for a response, and talking with someone usually meant being in close proximity. That was how we communicated throughout all of human history, and our communicative tools evolved and adapted with us over that time. Then, in middle of the last century, everything changed and we are now faced with tools that change faster than we can keep track of them.

It is our further misfortune that the ubiquitous manifestations of that technology are tools that focus our attention on a computer interface regardless of the nature of the task we are trying to perform. Our focus here is how to change that interface so that we can intuitively focus on the task at hand, instead of the tool we are using.

3.2 A Model of Calm

The theories underlying this chapter are based in part on ideas discussed during an invited talk at Australia's National Information Communications Technology Research Centre of Excellence (NICTA) in Sydney Australia, on July 25th, 2013.

A position paper based on the Brown-Hitz Model will be presented on April 26th, during the Workshop Peripheral Interaction: Shaping the Research and Design Space at CHI 2014, in Toronto Canada.

The paper, entitled "A Model of Calm" is in press. It was co-authored by John N.A. Brown, Gerhard Leitner, Martin Hitz, and Andreu Català Mallofré.

In order to establish a common understanding of interaction as a foundation for the discussions put forward in this thesis, a brief review of models of interaction is offered. We propose the Brown-Hitz model, which introduces the well-understood concept that humans take in and react to information consciously, pre-consciously, and reflexively, and proposes that input and output devices could be developed to do the same.

3.2.1 The Human Action Cycle and other Models of Interaction

Norman proposed a 4-stage model of the Human Action Cycle: "Thus the full cycle of stages for a given interaction involves: executing the action; and evaluating the outcome." [125]. Variations and modifications have been generated by numerous authors in order to illustrate the interaction between humans and computerized systems.

Among other models of interaction, Abowd and Beale [2] provide a model that shows the single incoming and outgoing actions on both the "human" and "machine" sides of an "interface" that is labeled with both "input" and "output" (redrawn in Figure 3).

Information is translated four times in a clockwise direction. A user's *task* is translated into *input* through *articulation*. The *input* is then translated again by a *performance* into a *core* that can be understood by the *system*. The *core* is then translated into a *presentation* for *output* which is subject to *observation* by the *user*. The translations become the focus of this model, in order to enable formal analysis of interface-based issues.

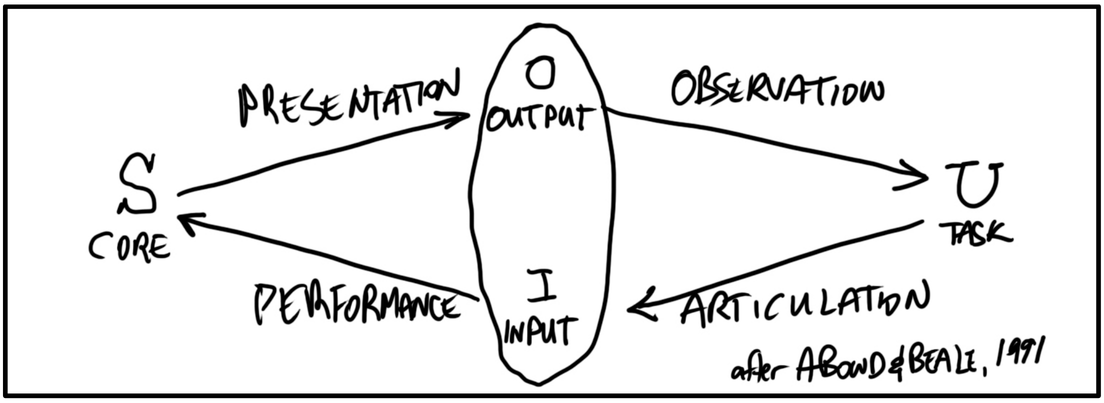


Figure 3: Abowd and Beale's translation-based model, Redrawn by brown to avoid © issues.

Mackenzie [117] simplifies the model, giving the reader three items on each side of the dotted line that represents the interface. The two directional items on either side of the line are now labeled in terms of computer use. The human's "motor responses" now exert actions on the computer's "controls", and the computer's "displays" feed into the human's "sensory stimulae". This is redrawn (with the components flipped horizontally in order to maintain a consistent direction across figures) in Figure 4.

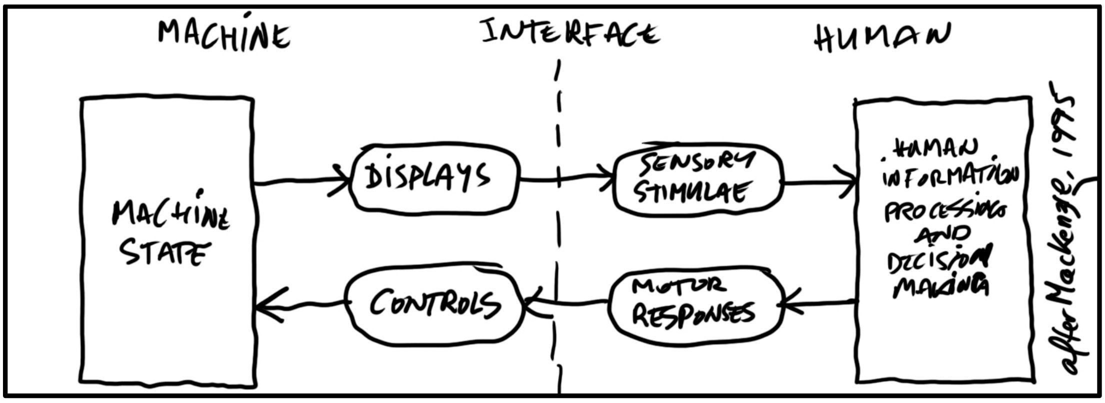


Figure 4: Mackenzie's simplified interaction model. Redrawn by brown to avoid © issues.

At APCHI, in 1998, Coomans and Achten introduced a more complex model, one that illustrates the processes between each action, labeling them with such descriptive terms as "thinking", "representing", "rendering", and "abstraction" (redrawn in Figure 5) [49].

Through naming, this model gives some implicit value to the differences between human and computer, telling us that the computer uses "processing" on "knowledge in internal digital representation", while the human conducts "rational thinking" using "knowledge" in an "internal mental representation".

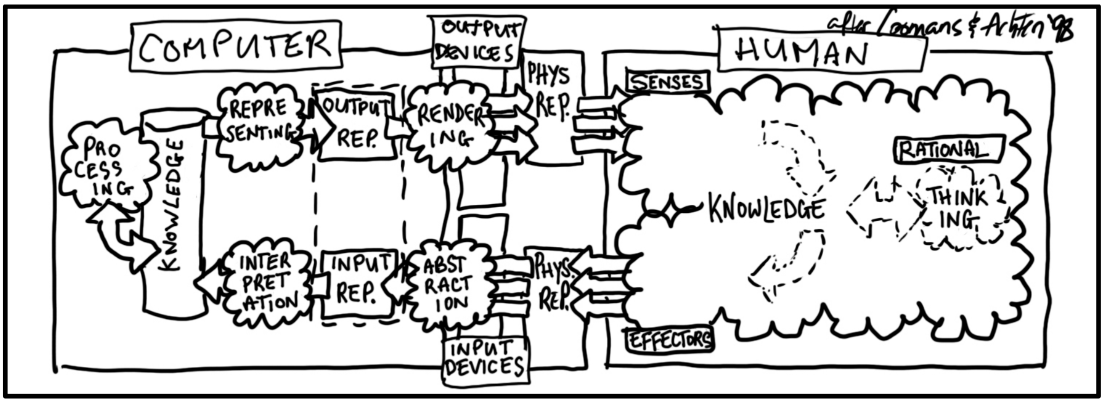


Figure 5: Coomans and Achten's model of interaction. Redrawn by Brown to avoid © issues.

The other matters of note are the way in which input devices are used to "abstract" human intent from a physical representation into something that can be interpreted by the machine, while the output device renders the machine's "output representation" into a "physical representation" that can be perceived by the human's "senses". This is of particular note when one is attempting to make HCI more human-centered, or calm, because the model shows separate arrows for different kinds of "physical representations" that have been rendered and presented to the human, and it shows separate arrows for the different kind of senses that might perceive these signals. What's more, it does the same for the human effectors, suggesting that there may be multiple separate channels of human output that could be driving computer input devices.

3.2.2 The Brown-Hitz Model of "Calm" Human-Computer Interaction

We propose a further modification of the Coonans and Achten model, to illustrate the means by which multitasking and peripheral interaction take place, thus pointing towards the HCI modifications necessary to enable Calm Technology (CT). Based on the theory of "Anthropology-Based Computing" (ABC), our model includes the "attentive", "pre-attentive", and "reflexive" information sensing and processing that take place in different parts of the brain and at different speeds. Figure 6 illustrates this natural aspect of human interaction with the world and suggests the possibility of deliberately-parallel input and output devices that would focus on one or the other of these sensing and processing modalities.

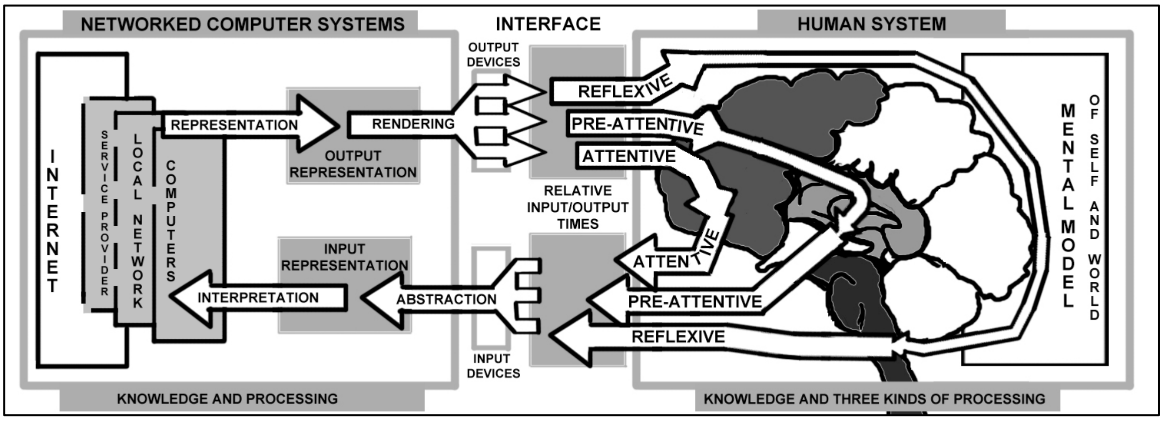


Figure 6: The Brown-Hitz model of "Calm" human-computer interaction. Note that incoming sensory information is processed in three different styles by three different parts of the brain.

3.3 A Measure of Calm

The theories underlying this chapter are based in part on ideas discussed during two invited talks. The first was at Australia's National Information Communications Technology Research Centre of Excellence (NICTA) in Sydney Australia, on July 25th, 2012.

The next invited talk was organised by the Malta section of the IEEE, and took place at Villa Bighi, the headquarters of the Malta Council of Science and Technology in Kalkara, Malta on October 12th, 2012.

A position paper based on the C.A.L.Matrix will be presented on April 26th, during the Workshop Peripheral Interaction: Shaping the Research and Design Space at CHI 2014, in Toronto Canada.

The paper, entitled "A Measure of Calm" is in press. It was co-authored by John N.A. Brown, P.S. Bayerl, Anton Fercher, Gerhard Leitner, Andreu Català Mallofré, and Martin Hitz.

Many researchers have tried to develop metrics of interaction [135], but the concept of Calm Technology (CT) is often misunderstood as a means of calming the user, as at The Stanford Calming Technology Lab (calmingtechnology.org). Others take it to mean the automation of decision processes as discussed in Makonin et al. [118], Olaru et al. [126] and

Stavropoulos et al. [143]. Such techniques have been criticized for failing to achieve Weiser's vision, and the criticism led to a call to abandon the concept in favour of "engaging rather than calming people" [138]. While those critiques may be valid, it is also possible to accurately focus on Weiser's intent, as discussed by Streitz and Nixon with the term "the disappearing computer" [145].

3.3.1 Introducing a Framework for a New Metric

The introduction of simple tools and metrics into well-studied areas of interaction has resulted in improvements in technological processes (consider the contrasting but pioneering work in motion study and efficiency conducted by Taylor [147] and the Gilbreths [78], and human performance [71, 72], Human-computer interaction [4] as well as in the reduction of human-error-related incidents and accidents in hazardous situations [100]. It has worked in these areas, but will it be possible to build such a tool for measuring Weiser's Calm?

3.3.2 "…the act of proposing such measures…"

Frankly, it may not be possible to achieve a universally-accepted metric for CT, but there is certainly value in attempting to do so. Consider the following quote from Fenton and Pfleeger [68]:

> *"Even when it is not clear how we might measure an attribute, the act of proposing such measures will open a debate that leads to greater understanding."*

The authors describe a six-step procedure for the development of a measuring scale for complex qualities that cannot be measured directly. They offer the steps quoted in this list:

1) Identify an attribute of a real-world entity;

2) Identify empirical relations for that attribute;

3) Identify numerical relations corresponding to each empirical relation;

4) Define mapping from real-world entities to numbers;

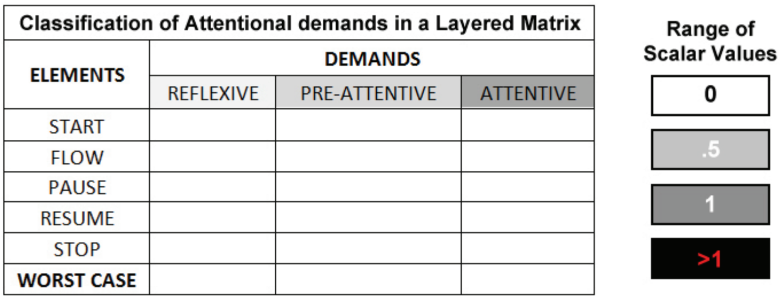
5) Check that numerical relations "preserve and are preserved by" empirical relations.

6) Combine the direct mappings from attribute to number (direct measures) in order to develop a model for indirect measurement.

At this stage in the development of our prototypical metric, we have achieved the first four steps by identifying three levels of attentional demands as potential attributes, and by giving them scalar values of “0”, “.5”, “1”, and “greater than one (>1)”. We fit the value of each attribute into a simple matrix in order to map attentional demands to those numbers and compare them within and across tools or tasks. The two final steps are not yet completed.

3.3.3 The Comparison of Attentional Levels Matrix (CALMatrix)

As a first usable step towards a quantitative measure of "Calm", we propose a matrix which matches the qualities of a task or tool to the reflexive, pre-attentive, and attentive demands involved in performing it. The prototype is shown in Figure 7. Defining a task in the matrix, one plots the type of attentive demand required during each element of performance. As mentioned above, at this stage, we apply only four scalar values: “0”, “.5”, “1”, and “>1” denoting that attentional level demand is either “not at all”, “partial”, “full” or, “over”. We hope that, in time, the matrix will become more sensitive. Perhaps the sensitivity will be increased by allowing the plotting of more intermediary partial values in order to more accurately reflect tasks whose demands are not all-consuming for any particular attentive level, or perhaps it will be through increasing the chronological sensitivity in order to better reflect task switching at any of the attentional levels.



Classification of Attentional demands in a Layered Matrix

ELEMENTS	DEMANDS		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START			
FLOW			
PAUSE			
RESUME			
STOP			
WORST CASE			

Range of Scalar Values

0

.5

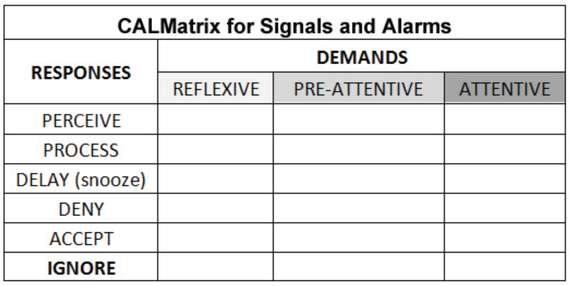
1

>1

Figure 7: CALMatrix for assessing tools and tasks.

Please note that we suggest that starting a task may be attentively different than performing it in "flow". We also suggest that there may be attentional differences in pausing, resuming and stopping a task. Finally, we insist that tools and tasks must be considered in light of the possibility of a “worst case” scenario, which might require fully-attentive processing at a moment’s notice.

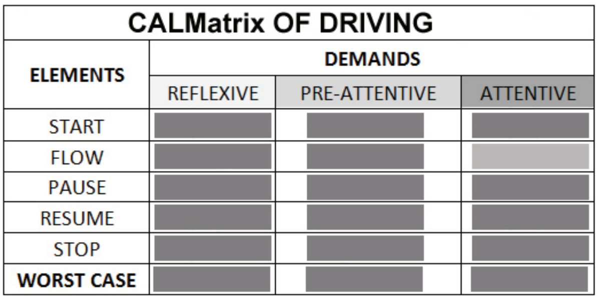
In order to measure how “Calm” a task or tool can be in the real world, one must consider not only other tasks that might happen at the same time, but also other disruptive events, such as interruptive signals and alarms. We are faced with a steady stream of alert, from the directed alarm that wakes us in the morning to the field of electronic crickets that chime text message alerts in a crowded train station, classroom or office. To account for this, we propose another prototypical CALMatrix, as illustrated in Figure 8



CALMatrix for Signals and Alarms			
RESPONSES	**DEMANDS**		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
PERCEIVE			
PROCESS			
DELAY (snooze)			
DENY			
ACCEPT			
IGNORE			

Figure 8: CALMatrix for interruptions, signals and alarms

Let us demonstrate now the use of a CALMatrix. We offer the example of "driving", which involves continuous processing that is both reflexive (haptic and optic) and pre-attentive (recognizing signs and markers, distances, movement patterns). Furthermore, because driving has potentially catastrophic consequences, the driver must always be ready to respond immediately with full attention, in a "worst case" scenario. One unfortunate element of driving is the false confidence that derives from performing a task successfully [116]. This gives the driver the illusion, especially when they are performing in “flow”, that they are completely aware of all conditions around them, and completely capable of dealing with whatever problems might arise [89]. As an illustration, please consider Figure 9, in which every cell in the matrix is filled in fully (denoting full demand) except for the cell denoting attentive processing during “flow”.

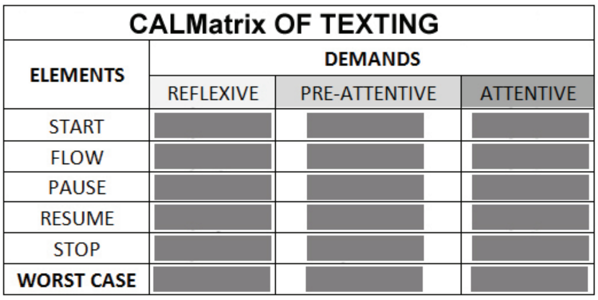


CALMatrix OF DRIVING			
ELEMENTS	**DEMANDS**		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START	■	■	■
FLOW	■	■	▒
PAUSE	■	■	■
RESUME	■	■	■
STOP	■	■	■
WORST CASE	■	■	■

Figure 9: CALMatrix for driving, illustrating partial attentive demand during “flow”.

That cell is filled in in halftone, denoting a partial demand. This reflects the fact that a driver need not pay full, conscious attention to their driving when it is going well. It is common for drivers to divert some part of their conscious and deliberate attention to other tasks at that time, such as conversation with a passenger, listening to a discussion or monologue on the sound system, or *staircase thoughts* about recent events. When in “flow” the possibility of simultaneously diverting one’s attention to other tasks is often mistaken for the ability to do so safely. This leads to “driver error” and has been linked to the majority of incidents, accidents, and fatalities.

Once matrices have been filled in, different tasks can then be compared to see if it is safe to perform them at the same time. In order to illustrate this, we offer Figure 10, a CALMatrix, showing the attentional demands during the exchange of text messages. These include, for example, reacting to visual and audio alerts (reflexive), recognizing photos, icons, customized alerts and the size of text blocks (pre-attentive), and thinking as one reads or composes a message (attentive).



CALMatrix OF TEXTING

ELEMENTS	**DEMANDS**		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START	■	■	■
FLOW	■	■	■
PAUSE	■	■	■
RESUME	■	■	■
STOP	■	■	■
WORST CASE	■	■	■

Figure 10: CALMatrix for texting, illustrating partial attentive demand during “flow”.

Please consider that, during an exchange of texts, both participants are focusing their attentive processing abilities as fully as they do during normal conversation. In fact, it has been suggested that, because of the lack of reflexive and pre-attentive signal and response during text messaging (as opposed to normal conversation) this kind of conversation actually makes more demands on the attentive processing abilities of both collocutors. In fact, it seems obvious, through observation and simple inductive reasoning that, while someone texting might believe that they are able to walk or drive or control a train at the same time, they are not. In fact, they are trusting that their peripheral attention will alert them to dangers, while inadvertently keeping those senses occupied with the reflexive and pre-attentive demands of the software and hardware.

In this way we can understand that it is not safe to combine driving and texting. Both demand attentive processing at all times and both are susceptible to the unwitting loss of peripheral perception. "Multitasking" might feel safe, but this false feeling is actually due to diminished attention to the periphery, not to diminished risk.

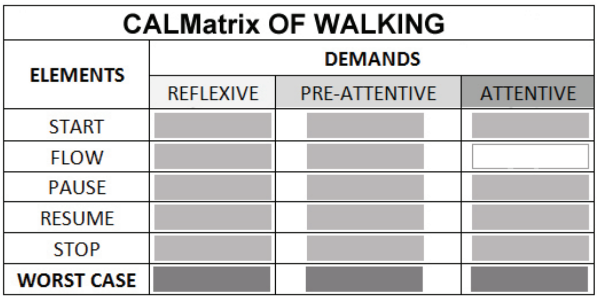
A final caveat: learning requires attentive processing, even when the task will eventually be performed reflexively, like walking or pre-attentively, like playing a familiar song on the guitar.

As will be discussed in future work, we plan to continue to develop our metric by following the guidelines mentioned earlier and conducting empirical trials.

3.3.4 Further Examples

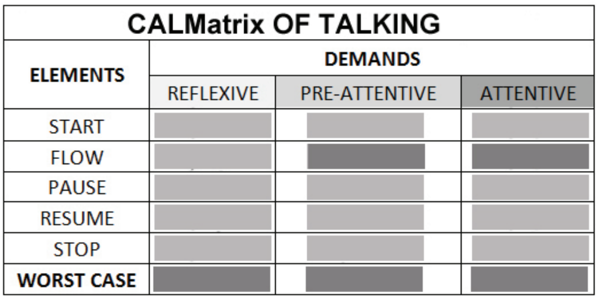
As further examples of the use of the CALMatrices, let us consider two common tasks, commonly performed together. Those of us who are able to walk tend to do so without a great deal of attention paid to the extremely complex interaction between the proprioceptors and related muscles that allow us to reflexively maintain our balance [50]. Despite our ignorance, the demand is constant, as reflected in Figure 11. There is also a constant pre-attentive demand to respond to changes in terrain, and to avoid walking into obstacles. When walking out in the open (as opposed to on a treadmill or private and secluded track), there is always the possibility that something will interfere with the reflexive and pre-attentive processes in a manner that will demand fully attentive response. Consider another example drawn from common day-to-day events. A crowd of pedestrians is waiting at a crosswalk for a break in traffic. If one person jaywalks, others may follow - if they pre-attentively register the movement without consciously checking to see if the light has changed or traffic has stopped. A shout from a concerned bystander, or the honking of a horn could then trigger the follower to pay attention to their situation, and so save them from an accident.

It is also possible that a walker in "flow" could follow a familiar path without noticing unexpected changes to the environment, and so step into something they might regret.



CALMatrix OF WALKING			
ELEMENTS	**DEMANDS**		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START	▨	▨	▨
FLOW	▨	▨	□
PAUSE	▨	▨	▨
RESUME	▨	▨	▨
STOP	▨	▨	▨
WORST CASE	■	■	■

Figure 11: CALMatrix for walking, illustrating how one might wander unthinkingly during "flow".

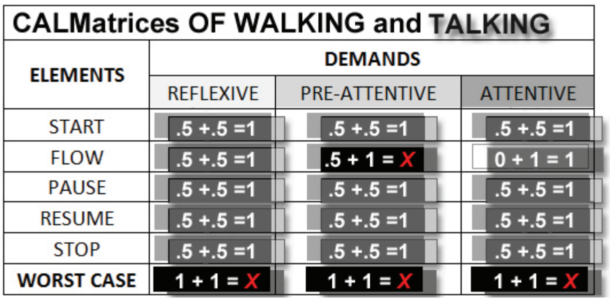


CALMatrix OF TALKING			
ELEMENTS	DEMANDS		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START	░	░	░
FLOW	░	▓	▓
PAUSE	░	░	░
RESUME	░	░	░
STOP	░	░	░
WORST CASE	▓	▓	▓

Figure 12: CALMatrix for talking.

Figure 12 provides a CALMatrix for talking. Please note that we do not consider, at this early stage, that full conscious attention is necessary for most conversations.

When one's conversation demands full conscious attention (illustrated here in the row marked "FLOW") it is also demanding on the pre-attentive processing that allows the recognition of facial expressions, posture and gestures, and provides the appropriate unconscious responses.



CALMatrices OF WALKING and TALKING			
ELEMENTS	DEMANDS		
	REFLEXIVE	PRE-ATTENTIVE	ATTENTIVE
START	.5 +.5 =1	.5 +.5 =1	.5 +.5 =1
FLOW	.5 +.5 =1	.5 + 1 = X	0 + 1 = 1
PAUSE	.5 +.5 =1	.5 +.5 =1	.5 +.5 =1
RESUME	.5 +.5 =1	.5 +.5 =1	.5 +.5 =1
STOP	.5 +.5 =1	.5 +.5 =1	.5 +.5 =1
WORST CASE	1 + 1 = X	1 + 1 = X	1 + 1 = X

Figure 13: Overlaid CALMatrices for walking and talking, with cumulative numerical values.

In order to consider walking and talking at the same time, we overlay the two previous CALMatrices as shown in Figure 13. In this overlay we see that, disregarding worst case scenarios, it is possible to walk and talk at the same time, except in the case that the conversation becomes so demanding that one must either stop walking for safety's sake because one's entire pre-attentive resources have been diverted to the conversation, or stop talking in order to proceed safely.

Again, the matrices presented here are prototypes and have not yet had the benefit of a few iterations of empirical testing, feedback, and improvement.

CHAPTER 4

Preparatory Studies

What follows are the published and as-yet-unpublished studies that allowed us to develop the formal tools we required in order to be able to set up and run the experiment that would answer our research question.

4.1 Hunting the S.N.A.R.K.

The chapter is based on

Brown, John N. A., Kaufmann, B., Bacher, F., Sourisse, C., & Hitz, M. (2013). "Oh, I Say, Jeeves!" A Calm Approach to Smart Home Input. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 265-274). Springer Berlin Heidelberg.

Our intent has been to create a system that would allow users to speak naturally with their Smart Home. Since the idea was first proposed formally by Weizenbaum in 1966, chatbots have gone in two directions. They either intended to fool humans into believing that the Turing test has been passed (here we get conversational patterns based on a misunderstanding of Rogerian psychotherapeutic questioning), or they are designed to elicit expected responses from a pre-determined database of choices (here we are subject to the prejudices of recommender systems, designed to believe that they know what the user wants to do better than the user does, herself). The prototypes of the pseudo chatbot used in our studies were built based on existing open-source engines, but were not be trained to "keep the conversation going at all costs" or to "help the user find her way to one of the available solutions". The intent of this pseudo-chatbox is to conversationally encourage the user to express their own ideas in a relaxed and natural manner, simulating the conversation two friendly humans might share when they are working together in a strict but informal chain of command. The version we propose here is intended to work the same way, but also to assist in the illusion that the user is dealing with a single holistic entity, rather than a network of embedded systems.

We propose to evaluate whether or not a speech and sound recognition software similar to the one described above can be made to work in an acoustically hostile environment, given the addition of a simple command protocol. This protocol is based on the triple-redundancy systems common to engineering we have already discussed, a truly user-centered perspective, and a hundred year old nonsense poem in which the captain tells his crew: "I tell you three times, it must be true".

4.1.1 A New Paradigm Part 1– The Bellman's Protocol

The underlying concept of Martial Arts training is that, through preparation, one can better deal with any situation that one can anticipate. I this manner, the difficulty of figuring out what to do and how to do it is chronologically displaced. The Chinese socio-cultural teachings that predate this attitude were addressed by Brown in a plenary session lecture at ICME 2012 [22]. In that talk, Brown proposed that, in the same way that this level of preparation prevents the martial artists from having to consider each new challenge individually, designers, programmers and engineers should be able to anticipate every use of their tools, and so should be able to make interaction less demanding at the moment of use.

As summarized by Nass and Moon [124], there is a lot of data proposing that humans accidentally and unconsciously interact with machines according to protocols established in human-human interaction. Surely this could be the key to making HCI intuitive and mindless. The user needs to have a familiar mental model to justify their input behaviors.

4.1.2 A New Paradigm Part 2 – The B.O.O.J.U.M.

A standard, generic lexicon and ontology will not allow the user to work within the personally-defined parameters of a mental model, requiring instead that the user(s) adapt themselves to their environment, in direct opposition to Weiser's intent.

The B.O.O.J.U.M. (Brown's Open Ontology for Joint User Management) allows each individual user to use their own preferred commands and names for each device, and is made up of two parts. The first is a device ontology based on formally recording the common values in each user's mental map of their home. This does not require any changes to the underlying ontology of the networked and embedded systems in the home; it is intended to assist the users, not the software. The second part of the B.O.O.J.U.M. is a 20-item command lexicon derived from a survey of 435 atomic use cases based on activities of daily living which will be discussed in 4.3.

The question becomes how to enable the computer to perceive the subtle signals that would reflect the mental model from which they are working, and the answer is surprising: concurrent signals conveying the same meaning improves accuracy of data transfer.

4.1.3 A New Paradigm Part 3 – The S.N.A.R.K. Circuit

We propose the *S.N.A.R.K. Circuit*, an early step towards a new paradigm of smart home interaction. Our goal is to provide a solution to the problem of using audio signals and voice commands in a noisy environment: a system that will wait unobtrusively to be called into service.

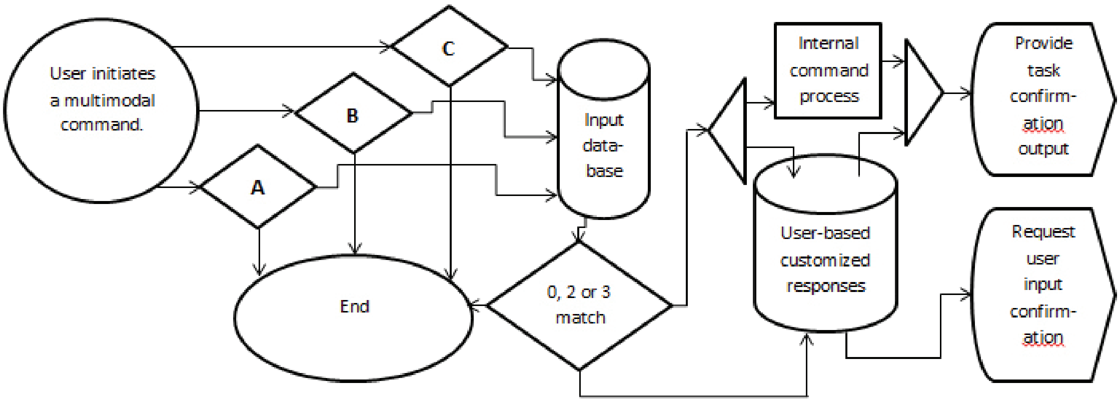

User initiates a multimodal command.
C
B
A
Input data-base
End
0, 2 or 3 match
Internal command process
User-based customized responses
Provide task confirm-ation output
Request user input confirm-ation

Figure 14: The S.N.A.R.K. Circuit: where any 3 recognized commands (A, B, C), detected within a small space of time, are compared to see if they hold the same meaning. If two of the recognized commands match, user confirmation is sought. If three match, the command is generated.

Figure 14 shows the *S.N.A.R.K. Circuit*, a "tell-me-three-times" command redundancy protocol or *Triple Modular Redundancy* [95] designed to fill the black box often assigned to filter noise from intentional command. Ideally, the trigger should be one that is easy to perform intentionally but difficult to perform accidentally. Conceptually, these parameters could be used to describe most naturally multimodal communication used by humans [63]; the combination of voice with the "separate symbolic vehicle" that we call gesture [120]. These can be simple actions such as using the space between one's thumb and forefinger to illustrate the size of an object while also describing it verbally. An ambiguous gesture can be easily misunderstood. For example, waving one's hands loosely in the air may mean several different things; from cheering in formal sign language to cooling burnt fingers, from saying hello to saying goodbye. Other gestures are less likely to occur by accident.

Our chosen example is the double clap. Clapping an uncounted number of times may be common, but clapping twice is well understood to be a means of getting attention from either a group or an individual. As a gesture, clapping twice is quite unique, in that it involves limited inverted movements coming immediately one after the other, and it is clearly delimited by a rapid start and equally sudden stop. While many people are familiar with the decades-old technology of double-clap sound recognition used as an on/off switch for electrical devices, this is not what we are proposing. We are proposing that the sound and the movement of the double-clap both be used as independent signals which can make up two of the three inputs recognizable and useable in our triple redundancy. This introduces one aspect of Calm Technology in that the user, intending to produce the noise of a double clap inadvertently produces the movement recognized as a separate signal. Inadvertent communication with a computerized system through natural human behavior is one of the key aspects of Calm Technology.

The trigger should also allow the system to distinguish known users from one another and from strangers. In "The Hunting of the Snark" Lewis Carroll wrote: "I tell you three times, it must be true!" We propose that a passive system could become active when triggered by three roughly simultaneous commands of equivalent meaning, delivered in different modalities. All three signals can be produced via the execution of a common human behavior for getting the attention of subordinates – the motion and sound of a double clap paired with the sound of a spoken name.

4.1.4 User Study

The *S.N.A.R.K. Circuit* relies on the detection of three different input signals. Brown et al. [26] have shown that the unique movement pattern of a double clap can be detected using the accelerometer in a standard smart phone. As mentioned above, in this study we have focused on the issues that were described as problems in the Sweet Home Project; sound detection in noisy environment.

To clarify this, Figure 15 offers the illustration of a single-modality version of the *S.N.A.R.K. Circuit*, detecting user intent with only two input signals (i.e. sound of clap and spoken name).

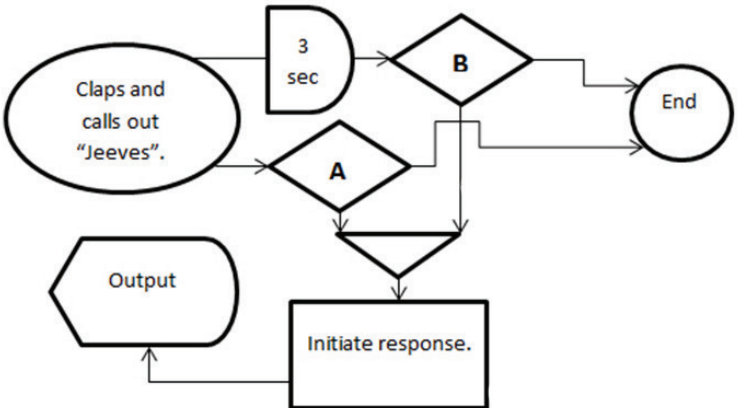


Figure 15: The Simplified S.N.A.R.K. Circuit: where two predefined commands (A – the sound of a double clap, and B – the sound of the spoken name "Jeeves"), detected in sequence, within three seconds, are recognized as a single multimodal command.

When the passively attentive system detects the sound of a double clap, it opens a three second window for recognizing the spoken word "Jeeves" (the name assigned to our system for this experiment). In future versions of this system, each user will assign an individually chosen name to the system.

4.1.5 Setup and Participants

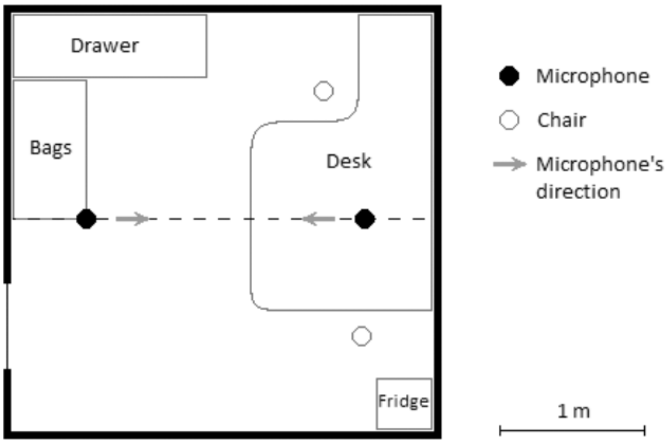


Figure 16: Layout of the experimental setting.

The experiment was performed in a 3m x 3m office arranged as shown in Figure 16. Two microphones (*Sennheiser BF812* and *Sennheiser e8155* connected to the PC via a *Line6 UX1* audio interface) were hidden along the medial axis of the room, at roughly 0.5 meters from the walls.

The number of microphones and their location reflects the work of Rouillard and Tarby [139]. The office is located in the administrative wing of a university, and so was surrounded by regular daily noises. On one side, there was an administration office and, on the other, a class room with computers. Noises of chairs moving or pieces of conversations could sometimes be heard, but the nature of our algorithm prevented that from becoming an issue. As will be explained in detail later, our system was constantly measuring noise levels, calculating the average over the ten most recent samples, and using that changing baseline as the comparative measure for recognizing a double clap. More precisely, a peak was defined as any noise over 120% of the background noise and two peaks occurring within 0.3-1.0 seconds were recognized as a double clap. This dynamic system is similar in nature to the interaction of humans who must adjust their volume as the volume around them changes.

10 men and 10 women, aged 19-28 (m=23, SD=2.635) from 8 different home countries, participated in our study. All participants said they were familiar with the use of a double-clap as a means of getting the attention of an individual or group.

4.1.6 Protocol

Ideally, the users of a future *S.N.A.R.K. Circuit*-based system should each assign names to the system. Each user could assign a general, all-purpose name with or without additional specific names to deal with specific situations. This second option would allow each user to experience the illusion of easily distinguishable interactive personas with whom to interact when seeking to accomplish specific tasks in the environment. This could allow both the user and the system to more easily identify situational context when interacting.

Consider, for example, the household inhabited by an engineer, a nurse, and their four year-old daughter. The engineer dislikes play-acting and prefers to deal with the single manufacturer's default persona for their smart home, while her husband prefers to deal with a smart home persona that is more like an *aide-de-camp* or servant. Their daughter secretly believes that there are dozens of people living in the walls of the house, some of whom are strict and only speak to her in harsh voices about safety rules, while others invite her to play and answer her by name when she calls.

At this stage of experimentation, we assigned a single name to the system for use by all participants: "Jeeves". We also used a single default automated response to successful attempts: "How may I help you?" If the system does not recognize that an attempt has happened, then it gives no response. For our current purposes, this was enough, and the exercise ended there.

When the full *S.N.A.R.K. Circuit* is implemented, the protocol is based on triple redundancy rather than double. To clarify, three recognized commands with the same meaning would initiate the automated response for successful attempts: “How may I help you?” In the case where only two recognized commands have the same meaning, the system would query the user for clarification: “I’m sorry, were you trying to get my attention?” Finally, if there are no meaningful matches between recognized commands, no response is initiated.

4.1.7 Data Collection

Participants began by “getting the attention” of the system named Jeeves. This meant performing an audible double-clap, and following it with a correctly-pronounced utterance of the assigned name. When the system identifies a double-clap, it then switches into a listening mode and stays in this mode for a few seconds until it identifies a sound as matching a word in the database, or until the timeout is reached. In both cases, the system then goes back into waiting mode until it identifies another double clap.

Participants did not have to orient their claps or their voice towards any specific target in the office. Initially, the participant enters the room, closes the door, and then walks around the office for a few seconds. Double-claps and word utterance were performed just after the entrance into the office, and after having walked around for a few seconds. An observer made note of *false positive*s and *false negative*s.

4.1.8 Software

Like Rouillard and Tarby [139], the recognition software was implemented in C# with the Microsoft System.Speech.Recognition library. This library allows direct access to Window's speech recognition engine. In order to detect double claps, the software periodically calculates the average noise level, using the AudioLevelUpdated event of the created speechRecognitionEngine object. Every time the signal level exceeds 120% of the average noise level of the last ten samples, the software detects a peek. If there are exactly two noise peeks within 0.3s - 1s, the software classifies them as a double clap. Every time a double clap is detected the speech recognition engine is activated for 3 seconds. If the word "Jeeves" is uttered within this period, the engine recognizes it and gives a response.

4.1.9 Results

For each part of the command recognition (i.e. the double clap recognition and the word utterance recognition) we classified the input data into four categories:

1) *true positive* (when the system is activated by a valid attempt),

2) *true negative* (when the system is not activated by an invalid attempt),

3) *false positive* (when the system is activated but no valid attempt took place), and

4) *false negative* (when the system is not activated but a valid attempt took place).

As illustrated in Figure 17, successful double-claps were performed and recognized 71% of the time. 10% of our events were *true negatives*, 2% were *false positives* and 17% were *false negatives*. Voice recognition was activated in 73% of our events (double-clap *true positives* + double-clap *false positives*). In other words, 73% of our original sample becomes 100% of the sample on which voice recognition is attempted. In this smaller pool, we found 83% *true positive*s, 7% *true negatives*, 3% *false positives*, and 7% *false negatives*.

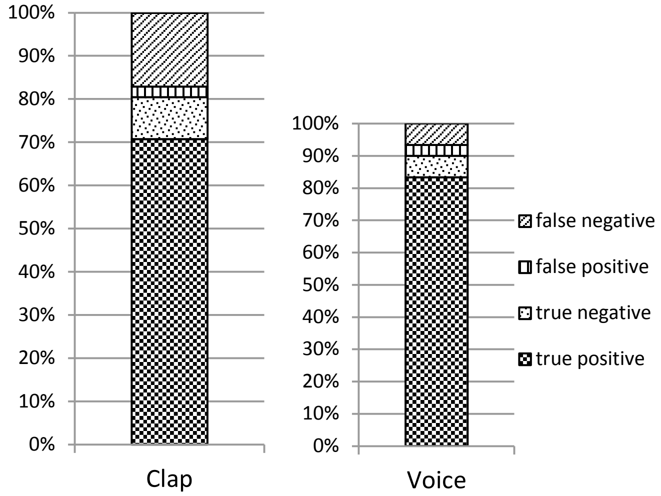


Figure 17: Clap and voice recognition. Voice recognition was only activated in cases when clap detection was either true or false positive (i.e. 73% of the double claps).

To calculate the overall system performance, we look at the number of times that the system performed correctly; identifying *positives* as *positives* and *negatives* as *negatives*. So we take all of the *true positives* and *true negatives* of voice recognition (i.e. 90% of the *true positive* clap recognition) and add it to the number of correctly-excluded double-claps (i.e. *true negative* clap recognition). This new total adds up to 76% of our original sample, giving us an overall system performance of 76%.

4.1.10 Discussion

Performance constraints were set for both double claps and word utterances to decide if they could be classified as *true* or *false positives* or as *true* or *false negatives*, and an observer was in place to label each performance as either a valid or invalid attempt. Double claps should be audible, and a short silence should be easily heard between each clap. Word utterances should be audible too, and the word "Jeeves" should be pronounced clearly. According to the observer, some participants performed very quiet double-claps the software simply did not recognize, while others did not wait long enough between the two hand-claps for our generalized standard (*true negative,* 10%). Two percent of the claps detected were actually background noises that deceived the system (*false positive*).

On the matter of speech recognition, some participants pronounced the word "Jeeves" incorrectly, producing instead either "Yeeves" or "Cheese". These mispronunciations were recognized 3% of the time (*false positive*), and were rejected 7% of the time (*true negative*). In this experiment, for the sake of expediency, default settings were used for the thresholds for recognizing double-claps and spoken words. Even the name "Jeeves" was used as a default. The interaction strategy proposed by this study is intended for customized environments and, as such, should include user-chosen names and user-modeled double-clap signals. Customizing the clap recognition to suit user abilities or preferences would certainly have improved the results.

In the end, the numbers are low. System performance of 76% is not impressive when considered in light of laboratory results. On the other hand, let us consider these results in comparison to the original studies conducted in Grenoble and discussed in some detail in 2.2.2. Fleury et al. performed a similar experiment in 2008 with 13 participants [73]. Their system attempted to recognize four types of deliberate noises and three types of scripted speech after extensive modeling. Their sounds had success rates of 81.25%, 100%, 42.86%, and 76.19%. This gives them an average sound recognition rate of 75.08%, slightly better than our sound recognition rate of 71%. Their different vocalizations were correctly recognized 30.43%, 83.44%, and 29.85% of the time, for an average of 47.91%. Our vocalizations were correctly recognized 83% of the time, in line with their best results and much better than their average.

Vacher et al. reported similar numbers in 2009, again using trained word recognition to attempt to recognize phrases associated with distress [148]. Their first attempt used a series of predetermined phrases in a silent setting (n=10) and reported an overall success rate of 84.4%. When they repeated the experiment (n=4) in a "noisy" environment similar to ours, their success rate dropped to 73%. Our overall result of 76% suggests that using the S.N.A.R.K. improved the result.

4.2 Pre-Attentive Gesture Recognition Using a Smartphone

The chapter is based on

Brown, J. N., Kaufmann, B., Huber, F. J., Pirolt, K. H., & Hitz, M. (2013). "… Language in Their Very Gesture" First Steps towards Calm Smart Home Input. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 256-264). Springer Berlin Heidelberg.

This experiment is another step towards the new paradigm of smart home interaction which will provide a solution to the problems stated above. As previously discussed, our eventual goal is a system that will wait unobtrusively to be called into service. Ideally, the trigger should be one that is easy to perform intentionally but difficult to perform accidentally, and one that can be used to distinguish known users from strangers and from one another.

We have proposed that a passive system could become active when triggered by three roughly simultaneous commands delivered in different modalities. All three signals can be produced via the execution of a common human behavior for getting the attention of subordinates – the double clap paired with a spoken name.

The previous section discussed our preliminary attempts to develop a system for recognizing that pair. The first question addressed in this study is whether or not the separate components of a multimodal communicative technique could theoretically be used as separate parts of a trigger as discussed above. Secondly, we ask if some commonly understood gesture exists (as per our Bellman's Protocol) and, if so, whether it can be used immediately, across cultural and linguistic barriers. Finally, we ask whether the signal communicated to the system could be extracted from secondary features of a deliberate human action.

We propose that the deliberate double clap can meet these conditions, and that it should be possible to automatically distinguish between common hand movements (such as waving or applause) and a deliberate double clap, as detected by the accelerometer in a smart phone. Smart phones have been used in many studies [105], [106], [103], [94], [6], but not with the intent that the signal should be generated incidentally during natural movement. This also addresses the issue of previous exposure to technology (PET) by allowing the human to use a common human behavior rather than behavior based on technological parameters or previous designs. In order to be certain that the phone's accelerometer is being used only incidentally, the entire device is mounted in an armband on the upper arm.

This is not to say that an arm-mounted smartphone will be the final form of the device. Raso et al. [134] mounted a smartphone in that manner for unobtrusive measurement of shoulder rehabilitation exercises and we have followed their example simply to be certain that moving the device remains an incidental action – an unconscious side effect of the attempt to create the desired double clap gesture.

4.2.1 Double Clap Recognition

For our smart phone application we used the LG Optimus 7 E900 and its built-in accelerometer. The application was implemented for the Windows Phone 7.5 framework. The accelerometer has 3-axes (X, Y, Z), and was set to 25Hz. It provides acceleration values normalized between -2.0 and +2.0.

We implemented a recognizer, which supports automatic segmentation to capture the double clap gesture. Performing a double clap is not only a commonplace gesture for getting someone's attention; it also has an easily-recognized pattern of accelerations and stops. In Figure 18 the raw accelerometer data are displayed as a continuous function where all axes are separated.

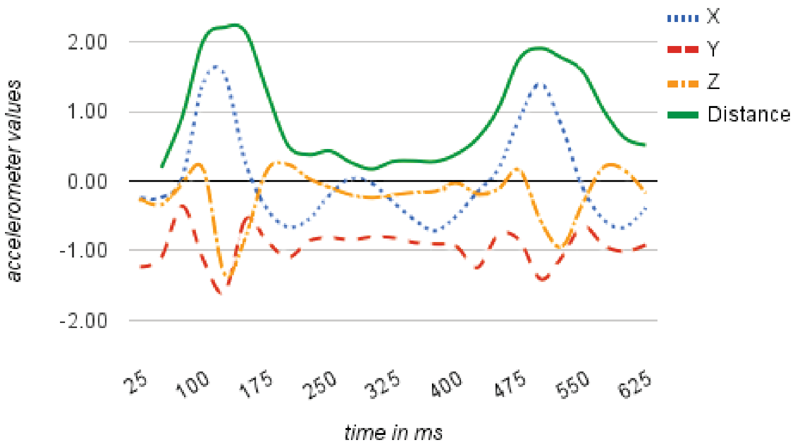


Figure 18: Accelerometer data.

The green line running above the others shows overall distance, as calculated using Equation 1. Significant regions are marked as follows. First, when the hands move towards one another, there is an increase in acceleration as shown in the first major increase in distance, the upwards slope of the solid green line. Acceleration increases until it suddenly plateaus and then changes direction, as shown in the drop in distance illustrated by the first downwards slope of the solid green line. The relatively horizontal region in the middle of the figure reflects the pause before the beginning of the second clap, as illustrated by the second major displacement in all of the lines.

$$\boldsymbol{D = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 + (z_k - z_{k-1})^2}} \qquad \textbf{(Eq. 1)}$$

Equation 1 gives us the difference between two consecutive accelerometer values, telling us the increase or decrease of acceleration between the two points. After every interval, given by the refresh rate of 25ms, the Euclidean distance to the previous collected accelerometer data will be calculated. If the average of the last three continuous distance values reaches a certain threshold, this is recognized as the end of a double-clap sequence. Once a possible end has been found, the algorithm looks back at the last 20 data entries (500ms) and examines the average of the distance between two consecutive axis values, to find both, how often the axis crosses the zero line, and any sudden changes in direction. Depending on the results, the recorded values and values collected in the evaluation phase were compared to one another.

4.2.2 Experiment

We conducted our experiment in four stages, with the assistance of 16 right-handed participants (3 females) between the ages of 23 and 31 (m = 27, SD = 2.2). Participants were first given a survey regarding their familiarity with the double clap as a deliberate signal.

Secondly the participants performed the double clap six times with the device mounted on their upper arm as described above, and six times without, in accordance with the method used by Kühnel et al. [106].

Accelerometer data was then collected while participant performed the double clap 10 times, without other motions between the repetitions. An observer took note of their performance, recording false positives and false negatives following the procedure described by Sousa Santos et al. [142].

The third stage was a deliberate attempt by each participant to confound the system. They were encouraged to try to elicit false-positives with any common or uncommon movements that, it seemed to them, might be mistaken for the acceleration pattern of a double clap.

The final stage of our experiment was another survey, asking what the participants thought of the double clap as an interactive method and soliciting their opinions on the difficulty of learning and performing the action.

4.2.3 Results

The first of our qualitative exercises was a resounding success: All participants confirmed that they were familiar with the double clap as a deliberate action and could generate it themselves. 56.3% confirmed that they had used a double clap to get someone's attention and 62.5% confirmed both that they have witnessed the double clap used to garner attention and that they could imagine using a double clap to initiate interaction with their computer.

Our second exercise was answered quantitatively. Our method measured successful double claps in 88% of the total number of trials. To be more specific, half of our participants succeeded in all of their attempts and two-thirds of the participants succeeded in over 90% of their attempts.

We excluded 2.5% of the overall number of double claps performed because they did not fall within the range of style, speed and/or noise level that was originally demonstrated.

Our third exercise resulted in the most entertaining portion of the data collection, as each participant tried to imagine and then execute some common or uncommon action which would be misread by our algorithm and mistaken for a double clap. Despite some wonderful performances, 75% of the participants were unable to deliberately generate a false positive. Two participants succeeded by vigorously shaking hands. Two others attempted to fool the system by doing the chicken dance. Only one of the dancers was successful.

In our final exercise, a post questionnaire, the participants told us how they felt about the ease or difficulty of using the double clap as a trigger for computer input.

Asked how easy or difficult it was to learn the double clap as demonstrated, 56.3% said that it was very easy and the rest described it as easy.

When we asked them how easy or difficult it was to perform the double clap, 56.3% reported that it was very easy, 37.5% described it as easy, and a single participant described the level of difficulty of the task as normal. No one rated learning or performing the task as either difficult or very difficult.

We wanted to know if participants would use a double clap gesture in order to interact with a computer. 50% responded with a simple yes, while 12.5% preferred I think so. One participant was uncertain, another chose I don't think so and 25% offered a simple no.

Those participants who did not like the idea of using a double clap as an interactive signal offered the following reasons:

"The gesture is too loud, it gets too much attention";

"Snapping the fingers would be easier", and;

"Clapping is not so intuitive."

Two said that clapping is for "getting the attention of animals" and four said that they would be more likely to use the device if it were "integrated into their clothes" or "in a watch", but "would not put it on every time" that they wanted to use it.

4.2.4 Conclusions

As mentioned above, in this study, we sought to demonstrate three key concepts.

The first is the possibility that a common gesture of human communication could be used as part of a multimodal trigger for getting the attention of a passive smart home system. The S.N.A.R.K. Circuit is such a trigger, and it has been proposed that the common attention-getting action of clapping one's hands twice and calling out a name could provide three separate signals. As discussed previously, the auditory components have already been tested, so we set out to test the possibility that the hand motions used to generate a double clap could be interpreted as computer input. We have shown that it can.

Secondly, we asked if some commonly understood gesture exists and can be used immediately, across cultural and linguistic barriers. Our attempt to find a gesture appropriate for our Bellman's protocol was mostly successful. All of our participants reported familiarity with the use of a double clap to get attention and agreed that it would be either easy or very easy to learn to perform it as demonstrated. However, as mentioned above, two participants had an intuitive mental model for the use of a double clap that differed from everyone else's.

Finally, we asked whether the signal communicated to the system could be extracted from secondary features of a deliberate human action. To answer that question we developed a straightforward accelerometer-based smart phone gesture recognition application, which could recognize hand movements. More specifically, we developed it to distinguish between general hand movements and the movement pattern of a double clap, allowing the recognition of this unique movement pattern as a deliberate signal.

The recognizer does not use any machine learning or other statistical probability methods. Instead it is implemented with a basic template matching algorithm using a distance equation to identify an increase or decrease in acceleration. User tests with 16 participants showed an accuracy of 88%. What's more deliberate attempts to deceive the system and induce a false positive met with a 75% failure rate, despite the simplistic mathematical method used. It seems that template matching is sufficient for the recognition of this rather unique gesture.

These results provide evidence that it is possible to find mental models in accordance with our Bellman's Protocol, and that the unconscious component(s) of natural human gestures can be used as deliberate components of multimodal commands, whether these components are redundancies for a simple trigger (as in the case of the S.N.A.R.K. Circuit) or whether they are, instead, either additive or stand-alone commands for more complex interactions.

4.3 AI-Based Recognition of Users and Use-Case-Based Commands with a Multilingual Population

This chapter is based on

an as-yet-unpublished study with Abdelhamid Bouchachia.

20 participants from 8 countries, speaking a total of 9 different first languages recorded a series of simple one-word commands derived from a list of 354 common activities in the home. Based on those recordings, post hoc application of a Gaussian mixtures models algorithm to a subset of our database of 2000 pronunciations of ten of our Use Case-based command words yielded user recognition rates of 94.5% and above.

4.3.1 The Use Cases

A list of 354 common actions in the home was compiled and sorted as use cases. Examples range from basic tasks such as turning on a light in the room you are in, to more complex tasks such as turning on lights in other rooms and even programming automated lights.

The use cases were then sorted into 94 categories based on attributes that could be fundamental in qualifying user intent. These 94 attribute-based categories were then grouped into 30 matching or contrasting sets, each constructed to be as broad as possible while still reflecting some specific aspect of how the use case could be perceived. Examples include simple pairs such as the contrasting relative locational descriptors “Remote” and “Proximate”. Other pairs included “Multi-modal” and “Uni-modal”; “Automatic” and “Directed”; and “Material/Organic” versus “Virtual/Digital”.

Other attributes formed conceptual triads, quartets or larger groups, such as whether an event would be “Unusual”, “Common”, or “Unique”; whether access should be “Individual”, “Family”, “Social (private group settings)”, “Civic (public group settings)” or “Open”; or whether a variable control setting is “Binary”, “Scalar”, “Initial”, “Radial”, “Final”, “Serial”, or “Cyclical”.

Finally, the original series of 435 use cases was tested to see how small a command vocabulary could be used in order to perform every task. In the end, it was decided that, along with basic identifiers for each device in the home, only 14 commands would be necessary: "On"; "Off"; "Open"; "Close"; "More"; "Less"; "Next"; "Previous"; "Select"; "Deselect"; "Set"; "Reset"; "Connect", and; "Disconnect".

4.3.2 The Use-Case-Based Ontology

The intent of organizing these attribute-based conceptual groupings was the facilitation of the design of a universal ontology for use-case-based navigation of a smart home tool set. This was initially proposed as a mathematical construct; a series of actions carried out individually or in varying combinations in order to perform a desired action in the smart home. This served the dual purpose of explicitly defining both the specific technological steps required to carry out an action, and the mental model of performing that action.

Imagine five layers of a cake. Layer one, the bottom layer is a list of every device (sensor, actuator, switch, etc...) in the home. On layer two you find the user(s)-designated name(s) of each location in the house that corresponds to any of the devices from layer one. Layer three is every simple relation that exists between, for example, devices and time. Layer four defines the more complex relations that exist between simple relations, and the topmost layer is the command lexicon that defines possible actions in the home.

To continue with the metaphor, the user slices into the cake, choosing a piece according to the mental model reflected by the lexicon at the top layer, without any concern for the basic devices and interrelationships that make it possible.

Let us break down an example. The resident of the home is leaving and issues the command "Leaving". This is a top level command that can be imagined as one of the cake decorations visible to the user. One layer below the perception of the user, the ontology defines the "Leaving" command as "All-Off" + "Auto Front Door Light" + "Motion Detectors" + "Burglar Alarm" + "Forward Communications".

At the next level down, the third layer, simple relations are defined. For example, "All Off" is defined as "turn off power to all non-essential devices", where the list of non-essential devices also has pre-defined values. "Auto Front Door Light" means that the motion sensors in a the "Front Door" space are set to trigger the light in the same space with a pre-defined sensitivity, and with predefined "delta t" for both triggering and subsiding.

Specific locations such as “Front Door” are defined at the next level down, the second layer, and specific devices are associated with them according to their actual distribution in the building. For example, several devices are filed as located near the front door, such as two sensors: s0001 and s0054; an actuator a0042; and a mechanical switch w0016.

The bottom layer is where the various identifiers or labels for each device are stored. A motion sensor would be stored here according to its serial number and brand name, with parallel cells identifying it according to the labeling strictures followed by the manufacturer, the contractor, the installation or maintenance team, and one or more of the users. Thus, we might find that the device referred to in the internal lexicon our ontology as “s0001” is also recognized in parallel lexicons as “Dual PIR/Microwave Motion Sensors”, as “DSC Products LC-171 with Pet Immunity”, and as “Motion Sensor Outside the Front Door”.

As we worked towards testing this use-case-based ontology, we also collected two thousand samples of spoken words, in order to test the Artificial Neural Network (ANN) we were planning eventually to use to recognize both our spoken commands and our users.

4.3.3 The Tests

Two Master’s students helped to collect sound files by recording volunteers as they spoke a 10-item subset of the 14 command words listed above. Twenty volunteers (10 female) from eight countries participated. They spoke an average of 3.6 languages each. Three participants were only bilingual and two others claimed proficiency in six languages. They ranged in age from 19 to 72 (m=25.45) and represented 9 different first languages.

Each participant recorded each command word ten times for a total of two thousand samples. The pools of each participant’s performance of each word were divided for testing. 60% was used to train the system with the remaining 40% held aside for testing the system.

The ANN used for this testing was not developed in the context of this thesis. In general terms, a Gaussian Mixtures Models algorithm was used to generate a model of each speaker, and each of the remaining utterances was then matched to each model. Sorted by command, accuracy ranged from 0,948 to 1,000, with a mean success rate of 97.8%. This compares very favorably to the results discussed in 2.2.2 as reported by Lecouteux, Vacher and Portet [109]. In a comparison of seven speech recognition systems subject to roughly 80 hours of training, they reported best results of 83.2%. Adding a Driven Decoding Algorithm (DDA) increased their success rate to 88.6%, 9.2% worse than ours.

4.4 The zAPP App: Exploration of Learning Complex Gestures with a Smartphone

The chapter is based on

"With a Wave of My Wand: Mobile Gaming with Speech and Orientation-Independent Gestures" an as-yet unpublished tech report, and

"zAPP: Gesture Learning and Transfer in an Informal Setting", an as-yet-unpublished study, both carried out in partnership with Mr. Istvan Fehérvári.

A prototype smartphone app for combined gesture and voice recognition was used in a study that took place at the annual gala ball at Alpen-Adria-Universität Klagenfurt to measure the ease with which passerby at a social event acquired the use of complex gestures.

4.4.1 Introduction

Now that we are in the era of Ubiquitous Computing [Weiser and Brown 1997], our gaming input devices are evolving beyond the mainframe and PC paradigms of the last century. Even the Nintendo Wii and Sony Playstation Move systems are, by their nature, as restricted to a single location as a game console from the 20th Century.

The smartphone has been used to make GUI, haptic and voice-based commands environmentally-independent, but has been restricted to 2-dimensional gestural inputs. Worse, specific orientation of the phone had to be maintained throughout the gesture. We propose a system for adding orientation-independent, fully 3-dimensional gestures to these other modalities and present a multimodal spell-casting game app in which player location and device orientation are not restricted.

4.4.2 Exposition

To achieve a complete multimodal interaction, the proposed system needs to recognize haptic inputs, accelerometer-based gestures and human speech simultaneously. In case of systems that receive input data continuously there is a need for advanced techniques to separate intended user inputs from noise. In this study we explicitly use touch to signal the

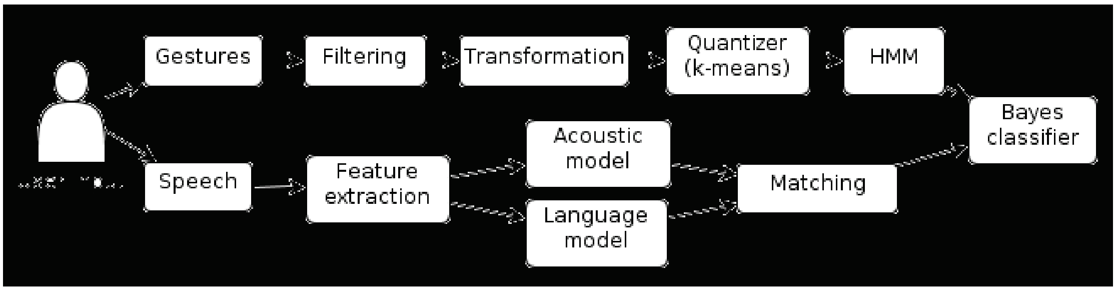


Figure 19: Overview of the recognition framework

start and endpoint of verbal and gestural commands from the user. The gesture and the speech recognition parts of the discussed system are completely separated, and their output will be fed into a Bayesian classifier. See Figure 19: Overview of the recognition framework for the architecture.

4.4.3 Gesture Recognition

Many accelerometer-based gesture recognition systems have already been developed using mobile devices. Although they provide very high detection accuracy, the gestures are usually simple symbols (geometric shapes or letters) that are restricted to a 2D plane in front of the user. In case of 3D gestures the detection accuracy could significantly drop if the user does not hold the device as it was trained phase [107].

To overcome these limitations we propose a system that follows the traditional machine learning approach, applying filters and hidden Markov models, but also uses the magnetometer and gyroscope; sensors which are becoming commonplace and yet remain largely unused. By filtering and fusing motion data from these sensors, we can obtain a momentary acceleration vector vi imposed by the user and a coarse orientation of the device Ri relative to a fixed world coordinate system in the form of a 3D rotation matrix. A transformation with the inverse of the rotation matrix is needed in order to obtain an orientation-independent acceleration vector. However, this method then requires that the user face the same cardinal direction in which the gesture was trained.

The solution is not complicated, but it has not been presented before. As shown in Equation 2, each vector needs to be further rotated around the vertical axis with the rotation matrix R_{yaw_i}, where yaw_i is the angle between magnetic north and the direction the device is facing.

$$\boldsymbol{v'}_i = \boldsymbol{v}_i \boldsymbol{R}_i^{-1}\, \boldsymbol{R}_{yaw_i} \qquad \textbf{(Eq.2)}$$

The sequence of such obtained v'_i vectors will then be used to create a left-to-right hidden Markov model for both training and recognition of the gestures. The open-source CMU Pocketsphinx library for word recognition was used in our pilot testing but not in the final trial, as explained below.

4.4.4 Pilot Testing the Device and the Bellman's Protocol

The first version of our app was introduced at the annual gala ball of the Alpen-Adria-Universität Klagenfurt. Promotional art is shown in Figure 20.

A gameplay station was arranged next to a dance floor, and partygoers in their tuxedos and gowns were invited to don wizard's hats and to challenge each other to play a game of "Rock-Paper-Scissors" using the zAPP [24, 66].

11 JAHRE
UNI
GALA
2012
zAPP
JNAB 2012
Developed by John NA Brown and István Fehérvári
Fakultät für Technische Wissenschaften an der Alpen-Adria-Universität Klagenfurt

Figure 20: Promotional Art from the Pilot Test

Two animated wizards appeared on a large television screen, each representing one player. The players and their corresponding wizards faced each other on a red carpet, like fencers before a duel. On the signal "en garde, pret, allez!" the two players cast one of three spells at each other – each with a different gesture.

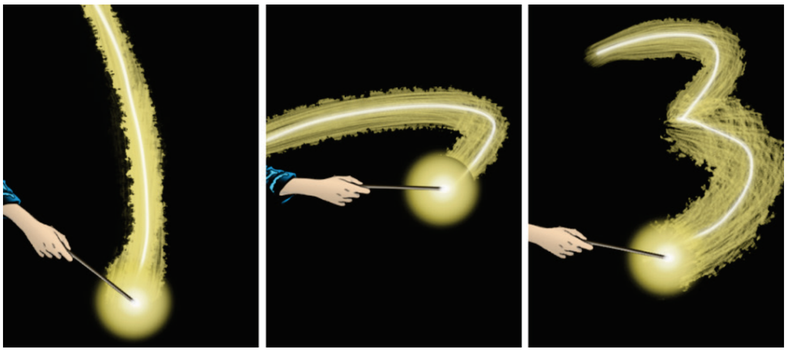


Figure 21: The Three Gestures

The three gestures were demonstrated by the experimenters or by other players and were visible as 2D illustrations, as shown in Figure 21. There was no familiarization period offered. Animations showed which wizard had won and which had lost.

The key here is that the players were not thinking that they were learning or using gestures as computer input. They were waving magic wands in order to play a familiar game.

4.4.5 Results and Conclusions

Over the course of five hours of open gameplay that evening, 649 gestures were attempted and 524 of them were successes. That is a success rate of 81%, in a party setting, cast by players who had undergone no familiarization. The location of the experiment next to the dance floor, and the boisterous nature of the event in general, prevented the use of voice recognition and still there was an 81% success rate.

With preparatory testing completed, the stage was set for trials of our integrated system.

CHAPTER 5

Interaction Unification in Distributed (Smart Home) Interfaces

We propose a system that inverts the common roles in HCI, forcing the machine to adapt its input and output requirements to suit the user. Our multimodal interaction is based on a modified version of the previously-described smartphone app that combines GUI, text, gestures and voice commands as a step towards natural human communication with a smart home.

5.1 Introduction and Setting

The Smart Home, like other networks of embedded systems, can appear to the user to be a single entity with multiple capabilities and interfaces. This would allow the resident to perceive “the house” holistically in the same way that a driver perceives "the car” or the average world citizen now perceives their hand-held communication device and camera as “my phone”.

The Living Lab smart environment laboratory at Alpen-Adria-Universität Klagenfurt has an open ceiling, exposing the lab to the regular office noises of an active research center housing four research groups, their technicians and their administrative support on the floor above. Furthermore, another laboratory and a set of offices have access to the lab through two doors, and use the lab as a passageway. Finally, the lab itself includes a functional kitchen and bathroom, both of which are at the disposal of all of the workers in all of those other rooms. This gives us a constantly-changing ebb and flow of background noise, including single and multiple voices, machinery sounds and sometimes a radio. It is our intent to state that such everyday noises do not interfere with the use of our system, even though noises and voices are crucial components of our input protocol.

5.2 Participants

32 volunteers (17 men and 15 women, ranging in age from 17 to 47, with a median age of 27 (8.29)) participated in a test of the two multimodal methodologies for interacting with our new, holistic and intuitive smartphone-based smart home interface. They were recorded and monitored unobtrusively from another room (Figure 22).

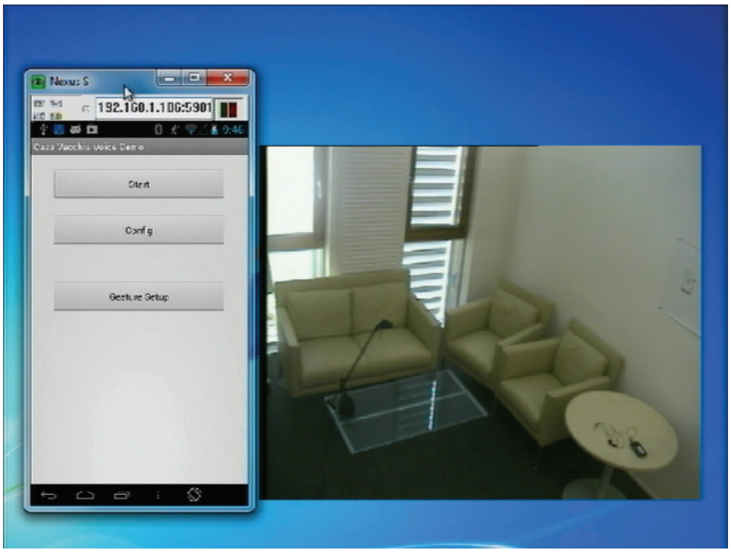


Figure 22: The living lab at Alpen-Adria-Universität Klagenfurt as seen through our hidden camera.

Our original cohort numbered 38. As in Cook et al. [46], some participants were removed from our study for not following the experimental protocol. Three recruits (two male and one female) were removed for refusing to follow directions. Two are researchers in HCI and decided to test the limitations of the system according to their own protocols. The other man failed to ignore our "noisy" environment. Instead of working around the people who passed through the room and walked past on the balcony above, he entered into two long conversations in the middle of the protocol.

Two others males were removed from the study due to their inability to understand the protocol as explained both verbally and in writing, in English. The data from a sixth participant was removed from the trial due to what seemed to be deliberately disruptive behavior by a particular group of people who were not involved in the experiment. Details are provided later in the discussion of the limitations of our study.

Each participant filled out pre- and post-trial questionnaires, supplying us with their demographic data and with qualitative feedback regarding their experience. The results compiled from these qualitative questionnaires are addressed in 5.8.2.

5.3 Familiarization

Here we build on the results of our zAPP app trials, based on the Bellman's Protocol, a reversal of the usual HCI paradigm requiring the adjustment of the user's skill set to suit the needs of the system. For that reason, it is expected that our system should be useful without familiarization. Participants were instructed in the basic methodology and given a written sheet describing the order of the tasks and the means of performing them. These descriptions were written in English. The participants were not given enough time or practice to familiarize themselves with the system [53]. As a result of this approach, three of our original 38 recruits (7.89%) were removed from the study due to an inability to understand the instructions as provided in English. When the choice was to either extend their exposure to familiarizing instruction or to remove them from the study, they were removed.

The participants were each engaged in a discussion of the history of personal servants, and told how the development of electric switches transferred the responsibility of performing some tasks away from the serving class and on to technology. In this way the participants were introduced to a mental model of their own homes populated by invisible servants waiting to be told to perform electronic tasks. This led into the suggestion that, here in the Living Lab, they could use an invisible *major domo* or *lead butler* to manage all of the invisible servants.

5.4 Testing Protocol

Each participant was presented with a list of nine simple household tasks that would usually require a mixture of absolute and scalar controls on four distributed mechanical devices. They were told that they would be asked to: 1) Turn on the light in front of them; 2) Turn off the light in front of them; 3) Turn on the Air Conditioner they cannot see; 4) Turn on the Radio that they can hear from the next room; 5) Turn off the radio that they can hear from the next room; 6) Open the blinds next to them; 7) Close the blinds next to them; 8) Open those same blinds a little more, and; 9) Close those same blinds a little more.

The C.A.S.A. T.E.V.A. app was then introduced and the participant was informed that they now had their own digital lead butler who would act as their personal intermediary to help get things done around the house. They were then asked to name their personal digital butler. Each participant was introduced to one of the two methods of interacting with their butler:

1) *m_g*, in which the participant focusses mainly on using gestures as the central modality, and

2) *m_v*, in which the perceived central modality is voice.

After performing the trial using one method, the participant was asked to start again using the other. The order was controlled, with 16 participants starting with each of the two manifestations.

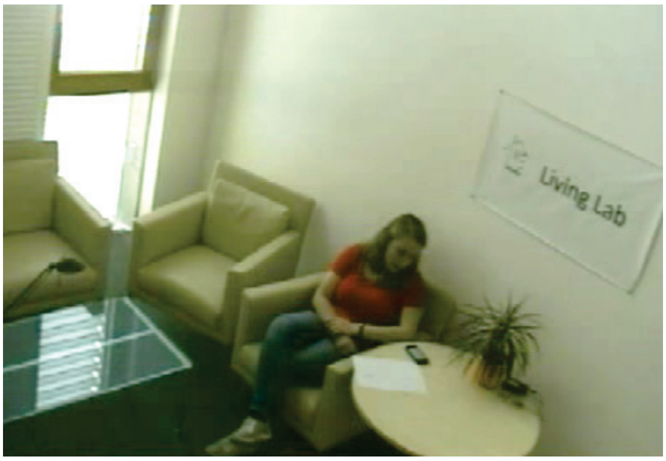


Figure 23: m_V: Voice-centered interaction.

5.5 The Voice-Centered Method

The first manifestation (*m_v*) is centered on spoken communication between the user and the digital assistant. Visual displays assist the user with their timing and with immediate feedback on the success of the different stages of their attempt.

An interaction scenario proceeds as follows: The participant is seated in an easy chair against one wall of the room, as in Figure 23, with the smartphone sitting on the table next to her. She double-claps her hands, in order to get the attention of the system, and she glances at the phone to see if it has reacted to that sound pattern among the background noises. If it has, it signals her with written text, asking which butler she wishes to contact. This response is only in writing because we did not want the participant to be annoyed by queries caused by false positives. In the next manifestation, signal recognition will be based around a full S.N.A.R.K. circuit as discussed earlier in this dissertation. That way, the S.N.A.R.K. will serve as a triple-redundancy fail-safe and reduce false positives and false negatives. This will be discussed in greater detail under Future Work in 6.4.

If the participant feels comfortable enough with the system (as was often the case in our experiment) she does not try to read the message. We see this in the majority of first attempts where the participant does not lean in or take hold of the phone in order to read the text. In cases where the system did not work on the first attempt, then the participant was more likely to try to read the message. During the first attempt, the participant either goes forward based on her own internal timing or she allows a general perception of a change on the screen to cue her to speak the name she has assigned to her personal butler before testing began.

"James", she says, and pauses again in her speech. It is important to note that these are not breaks in her speech, but pauses. She is speaking in the way that it has been modeled to her, with one- or two-second delays between each phrase. She is speaking clearly and slowly, as though to someone who is just learning the language. It is also important to note that, even though the participants were shown how to speak in this slow and deliberate manner, and even though many of them did so, many more simply spoke at their own natural speed.

Once the system has heard a recognized name, it queries the participant. In writing, a phrase appears on the screen: "How may I help you?" James also speaks to her in a slow and friendly, very human voice. "What would you like me to do?"

Referring to her list of tasks, the participant asks James to turn on the lights. The sheet of paper in her hand reminds her that she can use any normal English-language phrase that contains the keywords "turn on" and "light".

If James has understood, the desk lamp on the coffee table lights up. If at any time, James has not understood, he apologizes in both writing and voice, asking the participant to repeat her command.

5.6 The Gesture-Centered Method

The second manifestation (m_g) is centered around gestural interaction, but with required elements of speech- and GUI-based interaction as well. Again, written and aural feedback helps to keep the user aware of the stages of the process. An interaction scenario would proceed in the following manner.

5.6.1 Interaction Scenario

The participant is sitting or standing according to their preference. Several chose to alternate sitting and standing or to walk around while performing the tasks. One participant chose to lay down (Figure 24). Since the phone is already at hand, the participant presses the large button at the bottom of the ready screen labeled "Gesture". In the next iteration, if the button is released after less than 0,5 seconds, the signal will be discounted as an error. In the iteration that was tested, this results in a spoken error message: "I'm sorry, were you trying to get me to do something? I didn't understand. Would you please repeat that action?" At the same time, the written word "Sorry" appears on-screen.

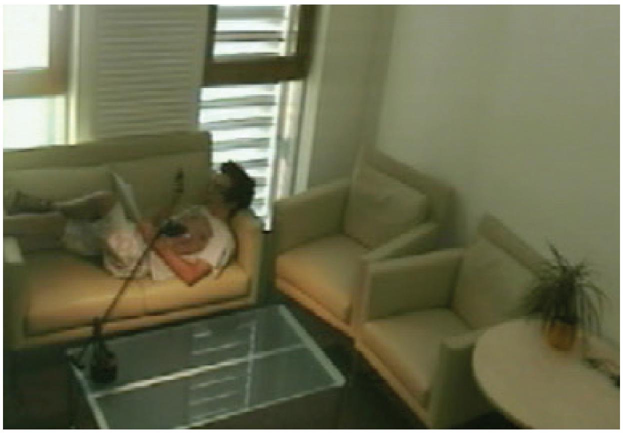


Figure 24: During gesture-centered interaction one participant chose to lie down.

If the button stays depressed, the following message appears on the screen: "Please name the device." After a brief pause, the participant names the device he wants to control. If the sound detected is not recognized as a known device or command word, the same error messages described earlier appear again, in both text and speech. At the same time, another message appears at the bottom of the screen: "I think you said: (x)", where x is replaced by whatever word was understood by the system. This extra written message is intended to provide additional feedback to the participant, so that they can correct their speech, if needed. This is intended as a polite reflection of the way in which a human collocutor would repeat a word that was not understood.

If the name has been understood, then there is no interruption in either writing or speech. The participant then waves the phone in one of three gestures, releasing the button once the movement is complete. If the gesture is not understood, the same error messages described earlier appear again, in both text and speech, including the message at the bottom of the screen.

If the gesture is understood, then the device and action are both correctly named in the message at the bottom of the screen. One example would be: "You said "light, off". As the text appears, the command is carried out. The message relating to the last action remains visible at the bottom of the screen, even as the rest of the screen returns to ready status.

5.6.2 The three gestures

The three gestures used in our trial are derived from the 19 basic commands that were a product of the review of 435 elemental use cases in the home. These use cases could be grouped into 94 different categories based on their function, modality, location or purpose. In developing our set of gestures, it proved possible to combine several of the 19 basic commands into a single concept: state change. As discussed in 2.2.1, a gesture need not reflect a word from any particular human language. Even though "state change" is not a common command in the English language, it does describe the intent of a number of different paired/binary commands such as: "on/off" or "open/close".

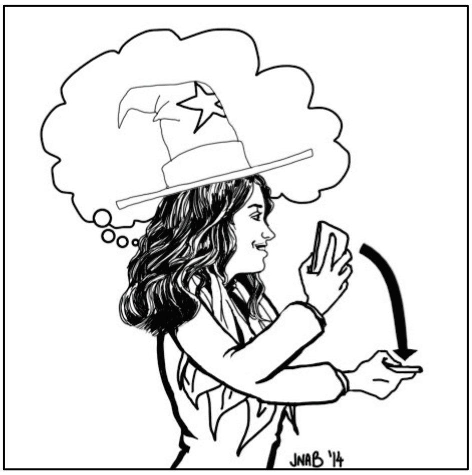


Figure 25: The mental model of using a magic wand, and the gesture for "state change"

This being the case, a single "state change" gesture could serve as well for opening or closing the blinds as it could for turning on or off the radio, the lights or even the air conditioner. The gesture itself was illustrated in a manner related to the gestural metaphor of a magic wand. As shown in Figure 25, one need only imagine oneself a wizard with a wand in place of the phone in one's hand and then drop the tip of the wand in the manner familiar across cultures.

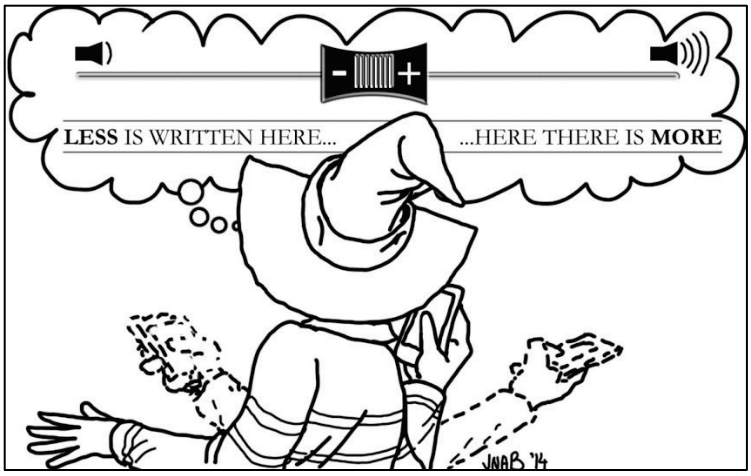


Figure 26: Mental model of left-to-right progression and the gestures for "less" and "more"

Two other gestures were necessary in order to carry out the last two of the nine tasks that made up our testing protocol. These are gestures for “more” and “less”. Here again we deviate from the work of Hurtienne et al., [91] as discussed in 2.2.1. Rather than basing the gestures on an assumed universally-understood spatial relationship between the person and their perception of the world, we chose to base our gestures on a global metaphor which is, itself, based on a cultural metaphor.

Volume sliders can be either vertical or horizontal. When horizontal, they increase from left to right, that is, in the same direction as word count in a written sentence – in most written languages. Even in cultures with languages traditionally written either right-to-left or vertically (Hebrew, Japanese, etc...), volume sliders universally work in the same direction.

On this basis, our gesture for “more” was to move as though increasing volume on a slider or word count in a written phrase, that is; a horizontal movement to the right. Our gesture for “less” was the opposite: a horizontal movement leftward. These gestures, and the mental models from which they are derived, are illustrated in Figure 26.

5.7 Outcome Measures

The new interaction paradigm, as expressed in our C.A.S.A. T.E.V.A. app, is an attempt to unify the control of a network of heterogeneous systems. The intent was to make transition from the control of one device to the control of another intuitive, and to make the transition from one modality to another seamless. Finally, we hoped to create in our participants the perception of the home as a single holistic entity. The conceptual basis for this has been described above. The attempted measurement of it is described below.

5.7.1 Intuitive transition between devices

The intuitive transition from controlling one device to controlling another was created by using a single perceptual and interactional ontology to control all devices. In this way, initiating a state change in one device required the participant to use the same behavior as when initiating a state change in any other device. This allowed the participant to work across devices following a single, holistic mental model.

This was tested by foregoing the usual protocol of familiarization. This lack of familiarization forced each participant to decide on their own how to transfer what they had learned about controlling one device to the task of controlling the other devices. If the participant did so on the first attempt, then intuitive transition had been successfully achieved.

Thus, the outcome measure for intuitive transfer is the ability of the participant to perform all nine of the tasks using the different systems embedded in our network. This will be discussed, and conclusions will be drawn, in Chapter 6.

5.7.2 Seamless transfer between modalities

The seamless transfer from one modality to another is, as discussed in 2.2.1, a naturally-occurring element of human communication, so long as the modalities are used in combination. Conversely, the use of a single modality is artificial and occurs only in response to a situation that prohibits natural communication.

Scuba diving requires single-modality communication (gestural interaction), as does use of a traditional telephone or two-way radio (vocal interaction). Our system attempts to recreate the rich multi-modal interaction that is intuitive to humans. This is done with the use of more than one modality, even when a single modality is the principle focus of interaction. This gives us two separate outcome measures for the seamless transition between modalities.

The first is the ability of the participant to transfer knowledge of how to perform a task using the gesture-centered manifestation of our app to performing the same task using a speech-centered manifestation, and vice versa.

The second outcome measure for the seamless transition between modalities is based on multimodality within each manifestation. More specifically, when using the speech-based manifestation of our app, the participant gets feedback aurally, but may also access additional information just by noticing that the screen layout has changed or by reading two different texts that appears on the screen.

The first one is a rephrasing of the aural message. The second one is a meta-level message, telling the participant "I think you said...". This gives the participant a deeper insight into why an interaction either worked or didn't, in much the same way that subtle facial or postural cues or word repetitions inform interlocutors during a human-human interaction. When using the gesture-centered manifestation of our app, feedback is also offered aurally, in writing, and as graphical change on the interface screen.

Thus, the ability to seamlessly accept feedback from multiple modalities is necessary in order to improve one's performance at a single task, and the outcome measure for this improvement is an increase in performance success as a task is repeated. For this reason, we evaluate performance in two ways.

The basic quantitative results given below were calculated on the basis that success at a first attempt was a pass and everything else was a failure. As mentioned above, this allowed us to test intuitiveness. However, we also collected comparative data considering success on first, second or third attempts. Improvement in performance as a result of feedback across modalities is the outcome measure for seamless transition between modalities within specific tasks. This is reflected in the comparison of performance in the first attempt to performance over the first three attempts. The conclusions we can draw on this basis are presented in 6.2.3.1.

5.7.3 Perception of the home as a single, holistic entity

The norm for smart home controls, as discussed in 2.1, is based on the concept that each device in the home has been grouped, in one way or another, into an ontological structure.

These ontological structures then become the templates for mental models that must be assumed by the user. For an example, let us consider the location of the "print" command in a word processing software. In order to print a document, you must first open the document, and then navigate to the control that allows you to choose the settings for the print job. In most standard software this navigation can be accomplished via GUI icon, via cascading drop-down menu or via hot key. To use any of these paths requires an understanding of where the "print" command sits within the ontology of the system. In other words, the user must learn how to navigate to the command that will enable them to attempt to achieve their goal of issuing the command to print.

The underlying function of the Bellman's Protocol and the B.O.O.J.U.M., is to turn this navigational demand on its head. As discussed in earlier chapters, our C.A.S.A. T.E.V.A. app is not the first to propose a virtual major domo, but it is to our knowledge the first working version to do so for the express purpose of eliminating the navigational demands of HCI.

In the C.A.S.A. T.E.V.A. system, one does not have to navigate from the entertainment system database to the lighting system database, or from one room's control menu to the control menu for another room. In our system it is only necessary to identify the device and the action. It is the responsibility of our major domo to carry out your intent. If the major domo does not understand your intent, it is their responsibility to clearly and politely ask for more information using customized phrases and terminology from the B.O.O.J.U.M..

The perception of the home as a single, holistic entity was measured using Anthropological methods for impartial observation and application of logical reasoning to the observed facts. Conclusions drawn from the observed performance will be described in 6.3.

5.8 Data Extraction and Analysis

In order to differentiate between true and false positives and between true and false negatives, performance was judged by system records and also by video observation.

For our experiment, observers followed the trials during the live performance via audio and video transmission to another room. After the trials, three observers rated each attempted performance into one of four categories as explained here:

- True positive, where an attempt was made to perform a task, the attempt was well-executed, and the result was successfully detected;

- True negative, where an attempt was made to perform a task, the attempt was not, for one reason or another, executed successfully and the attempt was not detected by the system;

- False positive, where either no attempt was made, or the attempt was poorly executed, and yet the system reacted as though an attempt had been executed successfully, and;

- False negative, where an attempt was well-executed and yet, for one reason or another, was not understood by the system.

In order to increase the likelihood of the accuracy of these categorizations, the video record of each trial was reviewed independently by three observers. Each was familiar with the protocol, with the tasks, with the setting and hardware, and with the device. Each observer was familiarized with this procedure in a pilot trial. Working at their own pace and in separate locations, each observer took descriptive notes for each performance and made a judgment – either assigning one of the four rankings for each attempt, or taking note of an inability to do so. After all videos had been judged by each observer, the three met and reviewed their results.

If two or three of the observers agreed on an outcome, the outcome was accepted as accurate. If there was no consensus at all on what had happened, the video was re-watched by all three observers together and discussed. This discussion continued until consensus was reached.

5.8.1 Quantitative results

Participants used both the gesture-centered method (m_g) and the voice-centered method (m_v) of multimodal interaction to perform 9 tasks. Descriptive analysis shows that task success averaged 55.95% (28.13%-68.75%) when performed with the voice-centered method, and 64.84% (34.38%-90.63%) when performed with the gesture-centered method. Let us examine these results more closely.

The 32 participants performed the nine tasks using both the voice-centered (m_v) and gesture-centered (m_g) multimodal interaction methods. The results were analyzed using two-tailed, paired student's t-Tests conducted at 95% confidence level ($p<= 0.05$) in all but three cases.

One participant did not perform task 1 ("turn on the light") and task 2 ("turn off the light"), and another did not perform task 7 ("close the blinds"). Since the number of performances to be incorporated into our calculations was not uniform for those tasks, the values for tasks 1, 2 and 7 were calculated using different sample sizes.

Looking at mean success rates based on first attempts by the total pool of participants, there is no significant difference between success rates (pass/fail ratios) based on the method used in seven out of the nine tasks (Table 1).

Table 1: Mean success rates across methods (m_v vs m_g)

Task	p-value
t1: Turn on the light	0,486
t2: Turn off the light	0,346
t3: Turn on the a/c	0,585
t4: Turn on the radio	0,007*
t5: Turn off the radio	0,007*
t6: Open the blinds	0,450
t7: Close the blinds	1,000
t8: Open the blinds more	0,217
t9: Open the blinds less	0,597

First attempts at these tasks using the gestural method of interaction both averaged 90.63%, well above average for mg (64.8%). Table 2 shows that order had no significant effect on the voice centered method.

Table 2: Mean success rates for m_v across order (m_v 1st vs m_g 1st)

Task	p-value
t1: Turn on the light	0,542
t2: Turn off the light	0,121
t3: Turn on the a/c	1,000
t4: Turn on the radio	0,481
t5: Turn off the radio	0,154
t6: Open the blinds	1,000
t7: Close the blinds	1,000
t8: Open the blinds more	0,492
t9: Open the blinds less	0,705

As seen in Table 3, the order of method use also had no significant effect on the success rate for the gesture-centered method.

Table 3: Mean success rates for m_g across order (m_v 1st vs m_g 1st)

Task	p-value
t1: Turn on the light	0,295
t2: Turn off the light	0,279
t3: Turn on the a/c	0,431
t4: Turn on the radio	0,559
t5: Turn off the radio	0,559
t6: Open the blinds	0,066
t7: Close the blinds	1,000
t8: Open the blinds more	0,729
t9: Open the blinds less	0,721

The mean success rate of each task by all participants is shown in Figure 27. Three tasks were attempted by only 31 participants. In all other cases task performance values were calculated from a base of 32 participants.

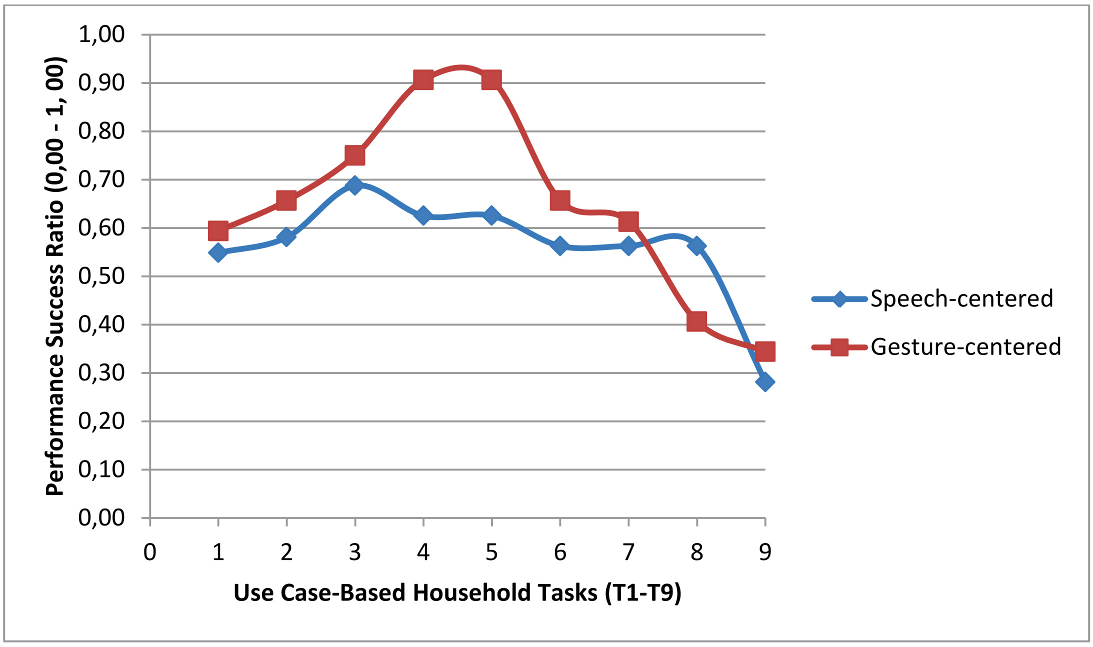


Figure 27: Mean success ratios according to interaction method.

Please note that, as mentioned above, the fourth and fifth tasks were performed with an uncharacteristic degree of success when compared with the other tasks. Please note also that the final task was performed with an uncharacteristically poor rate of success regardless of method, while the eighth task was almost as badly performed using the gesture-centered methodology.

The last two tasks (t8 "open the blinds more" and t9 "open the blinds less") were more conceptually difficult than the others. These scalar qualities of "more" and "less" required the use of new gestural commands, considerably different from the "state change" gesture used to perform the other tasks.

In terms of voice-centered interaction, the state change actions required that the voice recognizer understand a known action and a known device. When speaking scalar commands, the voice recognizer was required to interpret an additional command word (either "more" or "less"). It may be that these additional demands on the user and on the software, respectively, are responsible for the decreased success.

The variance between two sets of data can be compared using a Pearson's Chi-square test. Applied to our general pools of performance data (p<= 0.05), this allowed us to test the hypothesis that there would be no significant difference in success rate when using either interactive method.

Like the t-Tests, the Chi-square test also showed no significant difference between voice-centered and gesture-centered interaction (p= 1,000). It may be worth noting here again that the data was not sufficiently robust to allow for a Chi-Square to be performed on all first attempts by all participants.

As mentioned above, one of our thirty-two participants failed to attempt the first and second tasks, and another failed to attempt the seventh task. This left us with three holes in our data table. These holes could be worked around in performing the overall Student's t-Tests, but not in performing overall or specific Pearson's Chi-Square tests. Attempting to do so resulted in a "divide by zero error" and false results. Omitting the participants with the missing performances from the Pearson's Chi Square calculations enabled the generation of actual alpha values.

Thus, the alpha values for an overall comparison of the tasks by method had to be recalculated, excluding the participants who had missing data in their performances.

Table 4 displays the significance of the overall performance (minus the three individuals whose data sets were incomplete). Individual comparisons of each task, by method, are shown in the same table.

Table 4: **Difference in performance of the two methods (m_v vs m_g), in total and by task**

Task	p-value
All Tasks	1,000
t1: Turn on the light	0,997
t2: Turn off the light	0,999
t3: Turn on the a/c	1,000
t4: Turn on the radio	1,000
t5: Turn off the radio	1,000
t6: Open the blinds	1,000
t7: Close the blinds	0,999
t8: Open the blinds more	0,999
t9: Open the blinds less	1,000

As seen in Table 5, treating all nine tasks as a pool, there was no significant difference in performance based on order of use, for either interactive method (p= 1,000).

Table 5: **Difference in performance of each of the two methods (m_v vs m_g), by order.**

Order of performance	p-value
Mg 1st x Mg 2nd	1,000
Mv 1st x Mv 2nd	1,000

The heterogeneity of first-attempt results across interaction type, regardless of order of use is particularly interesting when one considers that the protocol did not limit the participants to first attempts. The implications of all results will be discussed in Chapter 6.

5.8.2 Qualitative results

We report on three different methods of qualitative data collection. Likert scales were used to gather descriptive information regarding our participants and to survey their opinion of the techniques used in the experiment. A system Usability Scale (SUS) was used to assess the usability of the system as a whole. Anthropological methods of observation and applied reasoning were used to assess performance and underlying meaning.

All thirty-nine of our original participants were surveyed before and after the experiment using standard Likert tests and the System Usability Scale (SUS). Furthermore, all were observed during the course of the experiment so that their performance and their behavior could be recorded. The data presented in this section reflects only the thirty-two participants who were, in the end, included in the quantitative data collection.

5.8.2.1 Likert scales: Perception of the system

Following the trial, each participant was asked a series of questions using a standard, five-choice Likert scale. There were sixteen questions presented. The first ten made up the SUS and will be discussed in 5.8.2.2. The eleventh question was a validation question, repeating a previous question to see if the participants were answering in a reliable fashion. All but one of the participants either repeated their previous answer or pointed out in writing that the question was a repeat. The exception was later removed from the pool because he seemed to have difficulty with the English-language instructions.

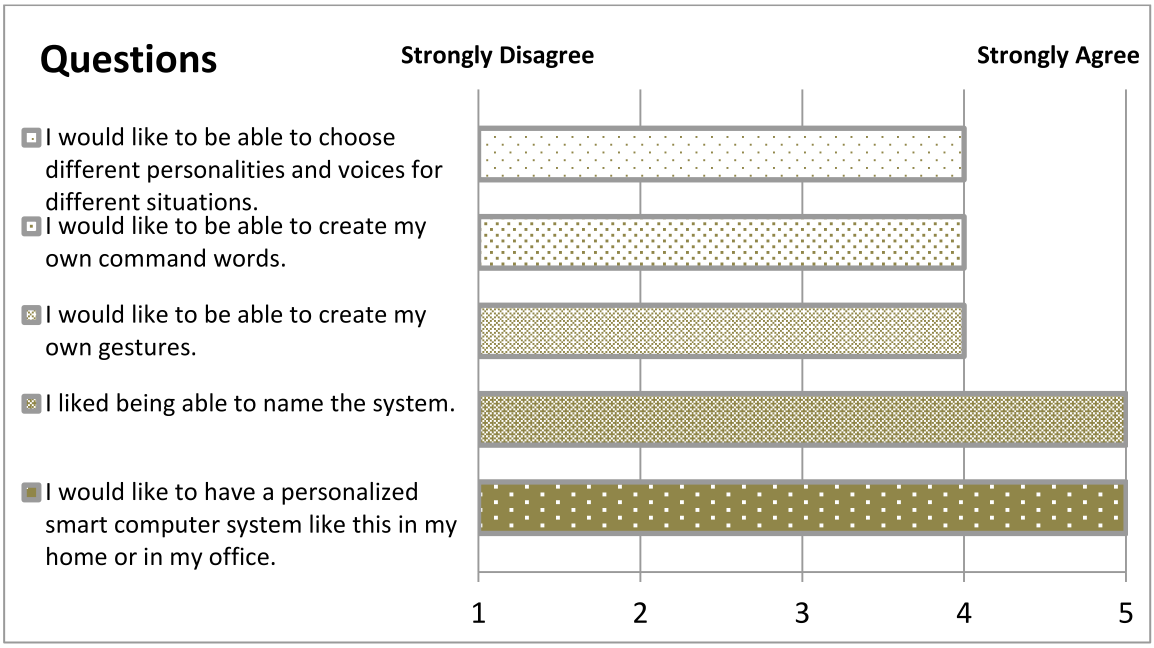


Figure 28: Participants' Opinion of C.A.S.A. T.E.V.A. Features (mode values)

The next five questions were a straightforward opinion survey regarding some of the features of the C.A.S.A. T.E.V.A. app. The questions and a summary of the answers are illustrated in Figure 28.

Two participants, one male and one female, both speaking German as a first language, gave the lowest possible answer (Strongly disagree) to "I liked being able to name the system". All other responses ranged between 2 (Disagree) and 5 (Strongly agree). Mode and median scores for that statement were both 5 (Strongly agree). The same was true for "I would like to have a personalized smart computer system like this in my home or on my office” (range: 2-5). The three other statements all had responses ranging from 2 to 5, and all had mean and mode values of 4 (agree).

5.8.2.2 The system usability scale

Developed by Brooke, the System Usability Scale (SUS) is a 10-item Likert questionnaire with set questions and a simple means of normalization [19]. The original paper has been cited over 1700 times. Even-numbered questions have negative implications and odd numbered questions have positive implications. In order to have consistent scoring, the even-numbered scores are inverted (subtracted from five). Then the answers are normalized to a four-point scale and then multiplied by 2.5 for expression in a range of 0-100. This expression makes the SUS values look like percentages, but they are not. An SUS value of 50% does not mean an accuracy of 50%. It means that the system being tested has a higher perceived usability than 50% of all such systems.

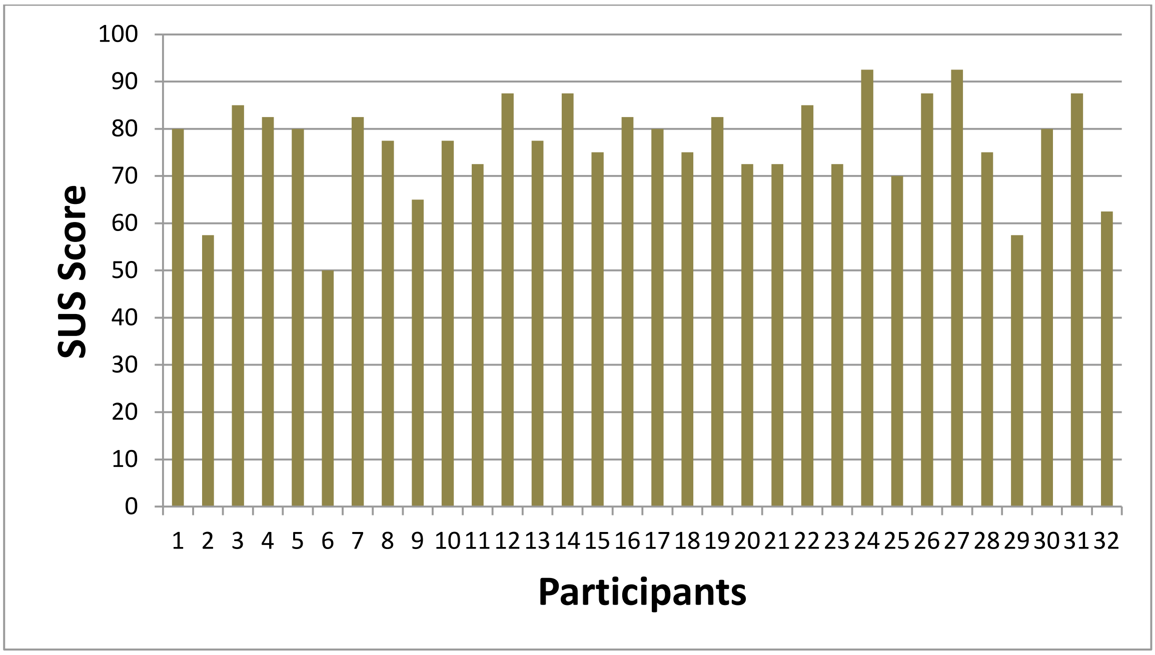


Figure 29: Participant evaluation of C.A.S.A. T.E.V.A. usability, according to the SUS.

Figure 29 illustrates each participant’s SUS rating of the C.A.S.A. T.E.V.A. app.

The participant ratings perceived usability ranged from 50 (1 participant) to 92.5 (2 participants) but trended towards the higher end, with a mean score of 77.03%, a median of 78.75% and a mode of 80%. The standard deviation was 10.15.

Based on these SUS scores, we can state that our participants, on average, felt that the C.A.S.A. T.E.V.A. application was more useable than 80% of all products tested in this area.

5.8.2.3 Anthropological methods and our conclusions

Likert scales are widely used, and the SUS has been used in over 500 studies since its inception. However popular these methods are in the field of computer science, the social sciences which depend upon qualitative data collection have largely turned to anthropological methods based on observation. This is because an opinion survey like a Likert scale can only collect the information that the user shares. Volunteered information is now considered to be inaccurate – whether consciously or unconsciously. In response to these concerns, we have introduced the use of anthropological methods in a few specific instances. These methods were specifically applied in order to answer three issues that will be discussed in details in 6.3: the intuitive transition between devices (6.3.1); the seamless transfer between modalities, both across and within methods (6.3.2), and; the perception of the home as a single holistic entity (6.3.3).

5.9 Discussion

Overall, the study reported in this chapter was a success. As seen above, the research question was answered and the null hypotheses were disproven. That said, the study could have been better designed and better performed. Some aspects of the study reported in this chapter are discussed here and conclusions are drawn. Aspects of the experiment with a greater bearing on the overall dissertation are discussed in Chapter 6.

5.9.1 Limitations of the Principal Study

Like all original formal experiments, the current study faced a number of situations where a choice had to be made regarding how best to deal with an obstacle, whether that obstacle was anticipated or not. Hereunder, we report on two such obstacles and our attempts to mitigate their effect on our experiment. Four additional limitations of our study are discussed in Chapter 6. We recognize that these limitations should be addressed in future studies.

5.9.1.1 King Midas' Ring

The ring we had designed and built to be part of this system did not work in time to be included in our study. The "Wireless Or Not" (W.O.N.) Ring was designed to allow the user to gesture with an empty hand, and have those gestures recognized as though the smartphone were in use according to our gesture-centered protocol (m_g). The sensor purchased for this purpose was easily modified to fit on a ring and was successfully pilot tested as a gesture recognition device. Unfortunately, the device as purchased does not have an "off switch". This meant that all gestures were always being recognized and were always being used as a single, steady stream of input into our S.N.A.R.K. system. The researcher was faced with the choice of adding an additional set of trigger gestures, modifying the S.N.A.R.K., or delaying the testing. Time constraints outside of our control eliminated the possibility of delays. In the end, it was decided that it would be easier to modify the software to deal only with two input streams than it would be to develop, test and finalize a natural and universal trigger gesture that would not provide an additional cognitive load for the participant.

Modifications have been discussed and additional testing of the modified ring will begin soon. It is our intent to include the ring in future tests of the smart home system.

5.9.1.2 Background Noise

This was a deliberate factor; the normal, day-to-day noises of a shared office and laboratory space. We wanted to show that background noises would not have a negative impact in terms of either generating false positive signals or in terms of obscuring user attempts. That said, on a few occasions passerby generated noises or actions that were deliberately disruptive. On several occasions, professors, researchers, students and staff engaged in loud conversations on the balcony overlooking the lab space, or while walking through the experimental setting. One staff member elected to play a radio loudly whenever our experiment was going on.

As mentioned earlier, one participant twice interrupted his protocol to enter into conversation with passerby who approached him. Given that this happened in only one of seventy-six trials – and then happened twice in succession- we elected to treat this as a conscious choice of the participant, and his data was removed from the pool.

The only other disruption caused by environmental conditions occurred when three professors carried on a very loud discussion with three students directly adjacent to a participant during her trial. The professors gave the distinct impression of trying to disrupt the experiment, talking much more loudly than usual and failing to notice the participant or the experimenter. During this disruptive behavior, first attempts at each task with m_v were all failures due to detection of speech from the disruptive conversation. The participant, showing signs of frustration, asked how she should proceed. She was encouraged to continue. In the end, she succeeded in performing five of the nine tasks successfully within three m_v attempts. Despite growing success the participant expressed her frustration verbally and non-verbally (by failing to complete the tasks, by self-interruption in synch with the disruptive conversation, and by walking off-camera). When her m_v trial ended the professors walked away. The participant's m_g trial ran smoothly with results within the normal range (44.4% success – almost exactly one standard deviation from the mean). The damage had been done. With a missed task in her m_v trial, and visible effect of frustration on her performance, her data was excluded from the pool.

With these two exceptions, the experiment was not impeded by the noises in the environment and the results we have described were achieved despite these potentially-disruptive behaviors. This may be due to a combination of the perseverance of the majority of our participants, the degree to which the general population has become inured to environmental distractions while interacting with their smartphones, or the robustness of the system. These would all be interesting matters for future research.

These and other recommendations for future work will be provided in the final chapter, along with conclusions, discussions, and a summary of the contributions.

CHAPTER 6

Conclusions

If Bardzell and Bardzell [15] were correct in their suggestion that Weiser's ideas of ubiquitous computing and calm technology have generated a bifurcated, rather than a holistic response; spurring engineers and technically-oriented scientists to pursue technological solutions, while driving psychologists and human-oriented scientists to pursue visionary new approaches, then this dissertation is an attempt to build a bridge between the two groups.

Rather than pursuing a purely technological solution to the problem of unifying the interaction with distributed interfaces throughout a smart home, we have attempted to create interactive technology that works based on how humans naturally perceive, process, and respond to environmental stimuli. Key to this idea is the understanding that human perception is based on fitting incomplete information into recognized and anticipated patterns. Trying to completely fill a recognizable pattern often leads to an "uncanny valley" experience where unconscious perception of "false notes" shatters the illusion [121, 77]. As any worthy magician or psychologist would agree, the secret is to provide just enough information so that the subject subconsciously fills in the missing pieces and creates their own, richer illusion.

6.1 Contributions

Our contributions to the fields of Smart Homes, HCI, and Calm Technology seem disparate when considered out of context, but set together in the right sequence they form a bridge between Weiser's concepts of Ubiquitous Computing and Calm Technology (Figure 30). This bridge allowed us to create among our participants the perception of a holistic and helpful Smart Home.

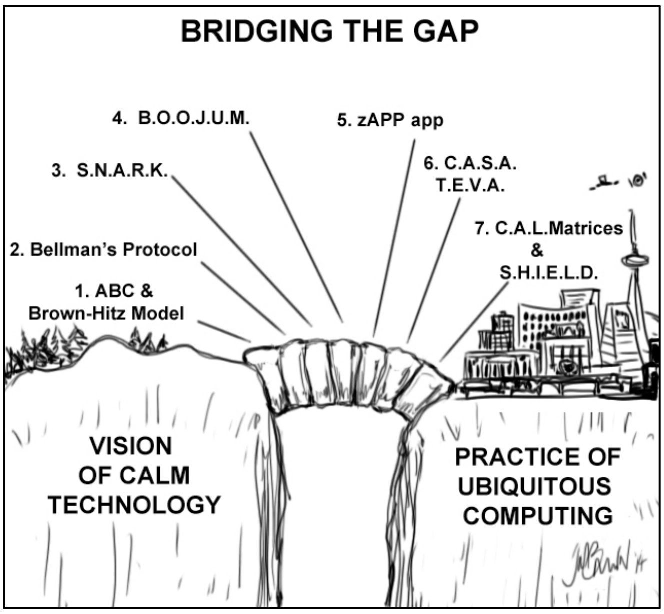


Figure 30: Bridging the gap between vision and practice

Contribution 1: Bardzell and Bardzell [15] called for a "cognitive update" of Weiser's model of HCI: "Calm Technology". Our first building block, the first stepping stone in our bridge, is one such cognitive update. As introduced in 3.1 and 3.2, we offer *Anthropology-Based Computing* (ABC) as a new conceptual model of HCI, and the *Brown-Hitz Model of "Calm" Interaction* (B-HM) as the illustrative model. ABC is quite literally a cognitive update to previous models of HCI in that it stresses the idea that human perception, cognition, and reaction must play the guiding role in the design and implementation of HCI. B-HM is the first such model to reflect that the human brain has more than a single cognitive processor, incorporating human peripheral sensory and cognitive abilities. A position paper on ABC was the keynote speech at IWANN 2012 [23], and a position paper on the B-HM has been accepted to the workshop on Peripheral Interaction at CHI 2014 [27].

Contribution 2: In a paper summarizing ten years of studying "...the breadth and depth of people's social responses to computers..." Nass and Moon [124] explained some of the factors that seem to trigger people to "mindlessly" treat a computerized interlocutor as though it were human. One factor that stands out in their summary is that, in interacting socially, the user is unconsciously applying a mental model of human-human interaction to their interaction with a machine. Nass and Moon are very concerned with the details of the triggers that cause this to happen and say that the next direction in their research is to try to precisely define all of the factors involved. Our next stepping stone is based on what they observed, rather than what they were trying to observe. *The Bellman's Protocol for Interaction with Smart Environments* provides a mental model for the user that is based on natural communication. The user should not worry about any of the underlying processes going on in the machine, or how to navigate between them – it is up to the designer to anticipate which tasks will be expected and to find a way to implement intuitive triggers for those tasks based on mental models that are already familiar to the user.

Contribution 3: Peripheral interaction is a necessary component of intuitive human communication [83]. Peripheral interaction with computers, however, often provides only subtle signals [12] that are difficult to detect under laboratory conditions and almost impossible to detect with any accuracy in the real world [13]. As a result, many previous attempts at detecting peripheral human input have largely been abandoned or relegated to "black box" design space. At best, attempts to capture these signals without extensive environmental equipment result in many false positives and negatives [73]. Our solution to this problem is the *S.N.A.R.K.* (Synchronize Natural Actions & React Knowledgeably), based on the triple redundancy protocols developed to ensure accurate communication with satellites [95]. The *S.N.A.R.K.* is based on two customized databases, and a small amount of automated reasoning. When active, the *S.N.A.R.K.* defines all incoming signals based on its user-generated input database. The definitions of all signals occurring within a short time-frame are compared. If two signals with an identical meaning are detected, the *S.N.A.R.K.* queries for confirmation. If three signals with an identical meaning are detected, the *S.N.A.R.K.* carries out the command. This application of an old solution changes the problem of multiple weak inputs into a strength – majority rules-based, triple-redundancy verification of intent.

Contribution 4: To date, the ontologies that underlie smart environments are proprietary and designed to suit a specific realm of associated devices [37, 45, 48, 55, 75,]. New devices cannot be added without customization [52, 104]. Furthermore, a different sort of customization allows a personal experience for the user [32, 38], but not for more than one user [33], and not in a manner that is easily adaptable [40, 57, 58, 127]. Generally, the systems are standardized [7] and require that the user(s) adapt themselves to their environment [9, 54], in direct opposition to Weiser's intent [25, 26]. We propose a smart home ontology based on allowing each individual user to use their own preferred commands and names for each device. The *B.O.O.J.U.M.* (Brown's Open Ontology for Joint User Management) is made up of two parts. The first is a device ontology based on formally recording the common values in each user's mental map of their home. This does not require any changes to the underlying ontology of the networked and embedded systems in the home; it is intended to assist the users, not the software. The second part of the *B.O.O.J.U.M.* is a 20-item command lexicon derived from a survey of 435 atomic use cases based on activities of daily living. A limited version of the *B.O.O.J.U.M.* was used in the major experiment described in Chapter 5. That version used six spoken commands, one sonification, and three gestural commands derived from the use case-based lexicon. The system was customized for each participant. They assigned names of their choice to the system and spoke using natural language, intuitively transferring their behavior for known tasks to other tasks according to their own mental model.

Contribution 5: It is "common knowledge" that gestures designed for HCI are difficult to learn [91, 96]. Despite that, gestures are a fundamental part of normal human communication [80]. That may be why the development of gesture-based interaction tools continues [43, 44, 69, 86, 90, 94, 97, 98, 101, 102, 111]. Due to smartphones, some touch-based gestures have quickly become almost universally-applied and accepted, but most in-air gestures continue to fall under the "common knowledge" mentioned above. What seems not to be generally understood by the designers of gestural interaction is that gestures are not normally and intuitively applied on their own. They are usually either a compliment to speech (as in natural conversation) [120] or a learned action that serves a specific meaning under a specific circumstance (as in the sign language used for scuba diving) [8]. We propose that the right mental model would carry with it meaningful gestures. Our challenge, then, became to map these known gestures to items from the list of use-case-derived action words we were proposing as commands. Secondly, we would have to develop an easy-to-use means of capturing the gestures as computer input. We decided that since smartphones are becoming ubiquitous, an app would be an ideal delivery system. The *zAPP App* allows users to interact with one another according to their own specific mental model of a magic wand; a cross-culturally recognized, hand-held device used for gestural commands. As a pilot test for the concept, we invited untrained and un-familiarized volunteers to use the system to play "Schere, Stein, Papier" ("Rock, Paper, Scissors"). This is described in Chapter 4. Results suggested that gestural interaction was easy to learn and use of it was based on a mental model that was already well-established. This became a fundamental principle for the design of our final experiment, as described in Chapter 5.

Contribution 6: The summative experiment of this dissertation combined the theories and technologies presented so far into a single system for holistic multimodal interaction with distributed interfaces in a smart environment. The *C.A.S.A. T.E.V.A.* app (Customizable Activation of Smart-home Appliances Through Enhanced Virtual Assistants) was tested with a sample of 32 participants. Using two multimodal interactive techniques, participants combined GUI, gestures, voice, and sonification to interact with a virtual major domo "who" helped them to use devices distributed throughout the Living Lab at Alpen-Adria-Universität Klagenfurt. The experiment and results are reported in detail in Chapter 5.

Contribution 7: The final stepping stone in the bridge between the theories of Calm Technology and the practice of Ubiquitous Computing is our *Measure of Calm.* Intended to be a practical, quantitative metric, at this time, it is only a series of prototypes. Our C.A.L.Matrix (Classification of Attentional demands in a Layered Matrix), which allows evaluation of any tool or task. We have also proposed a prototype CALMatrix for interruptive signals and alarms. Last in the series is a modification of the CALMatrix for use in identifying hazards and mitigating behaviors. The S.H.I.E.L.D. (Simple Hazard Identification through the Evaluation of Layered Displays) is intended for use by Human Factors Specialists. In developing the CALMatrix prototype, we have followed the steps set out by Fenton and Pfleeger [68] for the development of a quantitative tool. At the time of writing, we have followed four of the six steps. Our final two steps will be undertaken outside of this doctoral program. First there will have to be empirical trials to determine whether or not numerical relations "preserve and are preserved by" empirical relations. Then, if we pursue it, the final step will be to try to combine the direct measures of each attribute into a model for indirect measurement. What is important for our present purposes is to show that it is possible to cognitively transpose Weiser's philosophy of "Calm" into something quantifiable. Once this principle is accepted, be it through our attempts or through the work of others, the subjective quality of "Calm" will become objectively quantitative.

> *"As we learn to design calm technology, we will enrich not only our space of artifacts, but also our opportunities for being with other people. When our world is filled with interconnected, imbedded computers, calm technology will play a central role in a more humanly empowered twenty-first century." [152]*

6.2 Reflections and limitations

The work involved in creating this dissertation has spanned three years, seven countries, and two continents. What had seemed at first to be the purely technical challenge of developing new hardware and software was revealed to be a problem lacking the fundamental theoretical groundwork that would underlie such work. In attempting to develop the practical, neurophysiological and anthropological underpinnings of Weiser's "Calm", the solution to the original problem presented itself holistically; as a required combination of mental model and carefully-crafted illusion. Interacting with the S.N.A.R.K.

triggered the pre-attentive, pattern-recognizing processes in the minds of our participants to see a virtual butler trying to help them, even if he did not always succeed.

Unfortunately, some aspects of the S.N.A.R.K. were not executed as well as we would have liked. This hurt the illusion of helpfulness and reduced the quality of the experience.

6.2.1 The S.N.A.R.K. was not really a S.N.A.R.K.: triple redundancy

As discussed above, our protocol is based on the S.N.A.R.K. circuit: a triple redundancy input system based on command fail-safes developed for satellites more than a generation ago. Triple redundancy should give three possible interpretations of each perceived signal:

1) Three inputs of matching value is a clear signal of intent and the action should be carried out without further confirmation;

2) Two inputs of matching value is an unclear signal of intent. Therefore, before any action is carried out, the system should output a request for confirmation of intent. At this point it is possible to ask for a single, double or triple follow-up input to serve as confirmation;

3) When there are no inputs of matching value within the timeframe set for delineating the input series, all inputs should be treated as incomplete signals. In the normal course of events, the system continues to wait. If the system had already been triggered to expect a follow-up input – as in case 2), above – then it should either query to see if the expected follow-up is coming or alert the user to the fact that the allotted time interval has passed. This allows the user to easily rescind a failed attempt, simply by not following up.

Without the gesture-recognizing W.O.N. ring (as described in 5.9.1.1) the best case for input recognition could not occur. To put it simply, there was no way to give a perfect triply-redundant command. This means that our tests were conducted with a system that could only provide either failure or partial success. Given such a limitation, we are satisfied with our measured success ratios of .558 (m_v) and .649 (m_g).

6.2.2 Lack of full customization and language limitations

The intent of the Bellman's Protocol and the S.N.A.R.K. is to work with a totally user-derived lexicon, providing customized terminology for each activator and location in the house, as well as customized command words.

The underlying meaning of each term should be based on the use-case-derived command lexicon we have previously discussed, and the underlying meaning of each activator and location should be based on a state chart.

The viability of this totally personalized command system has been the subject of an as-yet unpublished study by Brown and Bouchachia using an Artificial Neural Network (ANN) to recognize a significant subset of the use-case derived commands and to recognize the voices of a variety of users. This ANN was not available in the year leading up to the study we are now discussing, and so could not be included.

Additionally, it was decided not to ask each participant to decide on their own command words and names for each device to be tested. To do so would be a natural part of installing a smart home system, but we were concerned that it would bias our participants as regards their perception of the system as a single, holistic entity.

The immediate result was that the system we tested was deprived of an ANN that has a greater than ninety percent success rate at recognizing users and commands. In place of that near-perfect recognition, we used the publically-available, on-line English-language speech recognition engine by Google. Subsequently, the system we tested had some trouble coping with the accents of our multinational participants.

6.2.3 Unfamiliar territory

Participants were shown how the device was used in both modalities. They were allowed to practice using the device, but they were not given time or opportunity to become familiar with the use of the entire command set. This was necessary so that we could measure whether or not they were able to intuitively interact with the system as a whole in a manner that seemed to them to be both holistic and logical, as per the original project description when it was initially conceived.

While this decision may have impacted negatively on our numbers, it also allowed us to draw a positive conclusion regarding the intuitiveness of the use of the system. Intuitive interaction and the perception of a logical and holistic smart home control system are discussed in 6.3.

6.2.3.1 A high standard of failure

The data used to derive our quantitative results were based solely on first attempts at each command. This is a high standard for a prototypical device, especially for testing in which familiarization was not provided (as discussed previously). That said, results were captured for multiple attempts. Participants had been directed to make no more than three failed attempts at each command before moving on to the next task.

The participants were significantly more successful when making one, two or three attempts than they were when only the first attempt is considered.

Success ratios for each task are grouped according to interaction method in Table 6.

Table 6: Success at 1st Attempt and by 3rd Attempt

Method	Task	Success Ratios	
		1st Attempt	1st , 2nd, 3rd Attempt
voice-centered	T1: lights on	0,548	1,000
	T2: lights off	0,581	0,903
	T3: ac	0,688	0,938
	T4: radio on	0,625	0,969
	T5: radio off	0,625	0,906
	T6: blinds open	0,563	0,906
	T7: blinds closed	0,563	0,844
	T8: open more	0,563	0,719
	T9: open less	0,281	0,656
gesture-centered	T1: lights on	0,594	0,938
	T2: lights off	0,656	0,938
	T3: ac	0,750	1,000
	T4: radio on	0,906	1,000
	T5: radio off	0,906	1,000
	T6: blinds open	0,656	0,938
	T7: blinds closed	0,613	0,871
	T8: open more	0,406	0,781
	T9: open less	0,344	0,781

Looking at the table above it is clear that average performance of each task across participants improved given a second and third attempt. Analysis revealed that the overall improvement was significant when success rates during first attempts are compared to success rates during first, second, and third attempts (as a group).

This comparison was made using a two-tailed within-group student's t-Test. The resultant alpha values, as shown in Table 7, reveal that the difference was statistically significant whether the performance was grouped according to method of interaction or generalized.

Table 7: Success Improved Over the 1st 3 Attempts

Comparing 1 attempt to 3	Alpha
t-Test of all tasks and methods	0,000
tTest of all tasks mv	0,000
tTest of all tasks mg	0,000

6.3 Conclusions

As discussed earlier, some evaluations of performance and underlying meaning were generated using anthropological methods of observation and deduction. These methods were specifically applied in order to judge true and false machine behavior during the trials, and to answer the three following issues: 1) the intuitive transition between devices; 2) the seamless transfer between modalities, both across and within methods, and; 3) the perception of the home as a single holistic entity.

6.3.1 Intuitive Transition between devices

The results presented immediately above show that, over the course of their first three attempts, the general pool of participants experienced significant improvement at all tasks. Please consider that, according to protocol, participants were not instructed in, or allowed to practice, the performance of every task. Participants were instead taught how to execute commands, but were only allowed to familiarize themselves with four tasks out of the nine.

In the case of m_g, the participants were each taught three gestural commands and were allowed to practice them on four tasks. In order to perform the five other tasks at all, each participant had to intuitively derive the relationship between the task they were asked to perform and the commands they had learned.

In the case of m_v, the participants were taught the syntax of the command structure and were given a task-order list which included the associated vocabulary. They were then encouraged to practice four of the nine tasks. They were not given enough time to familiarize themselves with the tasks.

The ability to improve one's performance must be based, at least in part, on the ability to intuitively generalize the tool use that was learned for the four training tasks, and apply it to the five other tasks. If this were not the case, the improvement would either be exclusive to the four learned tasks, or would at least be significantly less in the five tasks for which performance had to be intuitively derived. the participants not only intuitively transferred their new skills to the control of devices they had not learned how to use; their control of these devices improved over the course of their first three attempts – just as though they were familiarizing themselves with a learned task.

6.3.2 Seamless transfer between modalities

A second *sub rosa* goal of our research is to try to support the seamless transfer between modalities. Our trials were designed based on the broad theoretical concepts presented in the first section of this dissertation. The intent of the design has always been to force the computer to communicate in a more natural, more human manner. The result is multimodal communication designed to make use of the facts about natural human multimodality that have been discussed at length in Chapter 2.

While m_v is centered on voice, input is also generated via hand-clapping. While m_g is centered on the use of gestures, input also involves GUI button pushing and speech. In both cases, output is generated via speech, sound effect, GUI screen changes and two levels of written feedback, one of which provides a deeper feedback into the system's performance.

Given the multimodal nature of our interaction methods, "seamless transfer between modalities" could be taken to mean transferring between modalities while using a single interactional method, or when transferring from m_v to m_g (or vice versa).

6.3.2.1 Seamless transfer between modalities within methods

In order to successfully perform a single task using either of the multimodal methods described above, a participant had to be able to transfer from one modality to another, synchronizing their own gestural and vocal signals and interpreting multimodal responses. Given mean performance ratios of 55.95% (mv) or 64.84% (mg), we must conclude that the seamless transfer between modalities that was required for successful interaction did in fact take place at least 55.95% and 64.84% of the time.

6.3.2.2 Seamless transfer between modalities across methods

As was discussed in detail above, the C.A.S.A. T.E.V.A. app worked on the basis of the illusion of a virtual major domo. This illusion is intended to provide relief from the need to work with multiple heterogeneous systems and the resultant techno-stress. The participants were meant to perceive a single holistic system with which they could interact using either of our two methods. While it was always understood that participants might develop an immediate preference for one interaction method over the other, the intent was to provide the future user with the option of choosing between two different methods of interacting with the same major domo, based on preference or on changing environmental conditions.

If the transfer from one method to the other was seamless, then we would expect no significant difference in performance between the two, regardless of the order in which they were used. As our results have shown there was no significant difference in overall general performance between methods in seven out of the nine tasks (Figure 27). Order of use also showed no significant effect (Tables 2 and 3).

6.3.3 Perception of the home as a single holistic entity

The 32 participants who completed our study performed their nine tasks, once using our voice-centered methodology and once using our gesture-centered methodology. In each case they directed their command to the virtual major domo they had named. Every one of the users thanked the major domo at least once, calling it by name.

The participants were all working with a mental model that was centered upon the idea that the major domo – their personal invisible butler – would do the navigation for them. He would go from device to device, from control system to control system on their behalf in order to carry out their orders. The major domo navigated the ontology, but the participant did not have to.

Each participant performed their nine separate tasks, controlling four separate electronic devices in three separate locations, using a single mental model that carried over between interaction methodologies. Ipso facto, each participant was working with a mental model of their environment as a single, holistic entity.

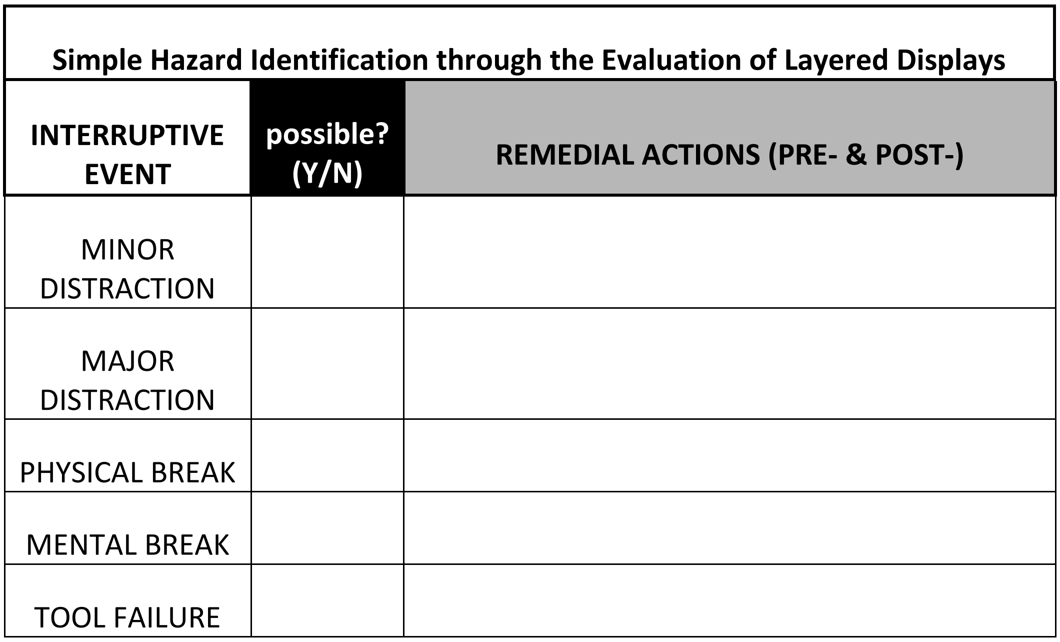
6.4 Future Work

This dissertation is an attempt, as expressed earlier, to build footings and lay the foundations for a new paradigm of HCI. It is our hope that Anthropology-Based Computing will be accepted as a means by which to finally bring about Weiser's vision of "Calm".

Our Brown-Hitz model of "Calm" HCI should be tested independently, as should our prototypical "C.A.L.Matrices", our S.N.A.R.K. and our B.O.O.J.U.M..

We ourselves plan to test a next generation of S.N.A.R.K. and B.O.O.J.U.M., part and parcel with a new generation of the C.A.S.A. T.E.V.A. app and a W.O.N. ring.

We are currently testing "Calm" ringtones and a "Calm" replacement for pop-up messages and trying to further develop our "C.A.L.Matrices".



Simple Hazard Identification through the Evaluation of Layered Displays

INTERRUPTIVE EVENT	possible? (Y/N)	REMEDIAL ACTIONS (PRE- & POST-)
MINOR DISTRACTION		
MAJOR DISTRACTION		
PHYSICAL BREAK		
MENTAL BREAK		
TOOL FAILURE		

Figure 31: Prototypical S.H.I.E.L.D. Risk Management Tool.

It is our hope that an accepted quantitative metric and the related improved understanding of peripheral interaction will not only improve the day-to-day experience of ubiquitous computing, but will lead to the acceptance of tools derived from the prototype shown in Figure 31 to help Ergonomists and Human Factors Specialists mitigate the dangers in high-risk fields. The point here is once again to identify qualities of individual tools or tasks in order to see what they require, but here the goal is to identify and mitigate hazards and risks rather than to promote "Calm". If there is a risk and there is no remedial action in place, then a task should not be undertaken. Once individual tasks have been evaluated, they can be layered with other simultaneous tasks. As in the previously-described C.A.L.M. tools, layering should reveal whether it is safe to perform any two or more tasks at once.

In their deeply reflective 2013 examination of the cognitive speculation underlying Weiser's conjoined visions of ubiquitous computing and calm technology, Barzdell and Bardzell [15] explain the rift in the work that has been done since. They posit that the technical work that could be done towards computational ubiquity was easily-envisaged and pursued, and that it has been updated in accordance with advances in the field. On the other hand, the visionary, human-centered aspects that should be based in psychology and anthropology have been set aside. They call for Weiser's vision to be updated "in a cognitive, rather than fantasy-based way". They offer hopeful guidelines for the updated vision:

> *"Steeped in a substantial mastery of ubicomp's empirical present (as Weiser was in his time), an updated vision will critically reimagine human experience in light of a present expert understanding of what a ubiquitous computing environment could be. This new vision will incorporate a holistic picture of what ubiquitous computing might look like, and it will relate it to an understanding of how human subjectivity itself might persist and change in this new reality, rendering visible previously hidden potential pitfalls and benefits alike. Such an understanding should destabilize present received assumptions about the "the user" and reopen the conceptualization of the user itself to both critical interrogation and empirical evaluation." [15]*

This dissertation is nothing more or less than an attempt to update Weiser's vision of "Calm". We have attempted to deal with both the visionary and technical sides of "Calm" and so to heal the rift in the original work that has been growing for the last two decades.

We have undertaken this healing from both sides of the wound, by giving the vision a rigorous scientific foundation in human cognition, and by giving the technical side working examples of a prototypical metric, and of an interface that uses a holistic and human-centered mental model to create the impression of "Calm".

References

1. Aaras, A., Dainoff, M., Ro, O. and Thoresen, M., 2002, Can a more neutral position of the forearm when operating a computer mouse reduce the pain level for VDU operators? International Journal of Industrial Ergonomics, 30, 307–324.
2. Abowd, G. D., & Beale, R. (1991). Users, systems and interfaces: A unifying framework for interaction. In HCI (Vol. 91, pp. 73-87).
3. Abowd, G. D., Mynatt, E. D., & Rodden, T. (2002). The human experience [of ubiquitous computing]. Pervasive Computing, IEEE, 1(1), 48-57.
4. Accot, J., and Zhai, S., "Beyond fitts' law: Models for trajectory-based HCI tasks," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1997, pp. 295-302.
5. Akgun, M., Cagiltay, K., & Zeyrek, D. (2010). The effect of apologetic error messages and mood states on computer users' self-appraisal of performance. Journal of Pragmatics, 42(9), 2430-2448.
6. Akl, A., Valaee, S.: Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, pp. 2270-2273, (2010)
7. Aldrich, F. (2003). Smart homes: Past, present and future. Inside the Smart Home. In Harper, R. (Ed.). (2003). Inside the smart home. Springer. pp.17-39.
8. Anastasiou, D. (2012). Gestures in assisted living environments. In Gesture and Sign Language in Human-Computer Interaction and Embodied Communication (pp. 1-12). Springer Berlin Heidelberg.
9. Angulo Bahón, C., Téllez Lara, R. A., & Universitat Politècnica de Catalunya. (2004). Distributed intelligence for smart home appliances : ESAII-RR-04-01
10. Atkinson, P. (2010). The Curious Case of the Kitchen Computer: Products and Non-Products in Design History. Journal of Design History, 23(2), 163-179.
11. Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science, 49(2), 324-330.
12. Bakker, S., van den Hoven, E., & Eggen, B. (2010, January). Design for the Periphery. In Proceedings of the Eurohaptics 2010 Symposium Haptic and Audio-Visual Stimuli: Enhancing Experiences and Interaction, Amsterdam, The Netherlands, July (Vol. 7, pp. 71-80).
13. Bakker, S., van den Hoven, E., Eggen, B. (2012). Acting by hand: Informing interaction design for the periphery of people's attention. Interacting with Computers, 24, 119–130.
14. Bakker, S., van den Hoven, E., Eggen, B., & Overbeeke, K. (2012). Exploring peripheral interaction design for primary school teachers. Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, 245-252.
15. Bardzell, J., & Bardzell, S. (2013). "A great and troubling beauty": cognitive speculation and ubiquitous computing. Personal and Ubiquitous Computing, 1-16.

16. Bell, G. A Congeries on the Computer-in-the-Home Market, Internal Memorandum of Digital Equipment Corporation dated 11 December 1969. Accession no. 102630372, Archives of the Computer History Museum.

17. Brandt, Å. Å., Samuelsson, K., Töoytäari, O., & Salminen, A. L. (2011). Activity and participation, quality of life and user satisfaction outcomes of environmental control systems and smart home technology: A systematic review. Disability and Rehabilitation-Assistive Technology, 6(3), 189.

18. Brod, C., Technostress: The Human Cost of the Computer Revolution. Addison-Wesley Reading^ eMA MA, 1984.

19. Brooke, J.: SUS: A "Quick and Dirty" Usability Scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London (1996)

20. Brown, J. N. A., (2004). A New Input Device: Comparison to Three Commercially Available Mouses (Master's Thesis, UNIVERSITY OF NEW BRUNSWICK).

21. Brown, J. N. A., Albert, W. J., & Croll, J. (2007). A new input device: comparison to three commercially available mouses. Ergonomics, 50(2), 208-227.

22. Brown, J. N. A. , "Expert Talk for Time Machine Session: Designing Calm Technology "… as Refreshing as Taking a Walk in the Woods"," 2012 IEEE International Conference on Multimedia and Expo, vol. 1, pp. 423, 2012.

23. Brown, J. N.A. "It's as Easy as ABC: Introducing Anthropology-Based Computing" In Advances in Computational Intelligence, pp. 1-16. Springer Berlin Heidelberg, 2013.

24. Brown, J.N.A. & Féhrevári, I. (2012) zAPP: Gesture Learning and Transfer in an Informal Setting. Unpublished report.

25. Brown, J. N. A., Kaufmann, B., Bacher, F., Sourisse, C., & Hitz, M. (2013). " Oh, I Say, Jeeves!" A Calm Approach to Smart Home Input. In Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data (pp. 265-274). Springer Berlin Heidelberg.

26. Brown, J. N. A., Kaufmann, B., Huber, F. J., Pirolt, K. H., & Hitz, M. (2013). "… Language in Their Very Gesture" First Steps towards Calm Smart Home Input. In Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data (pp. 256-264). Springer Berlin Heidelberg.

27. Brown, J. N. A., Gerhard Leitner, Martin Hitz and Andreu Català Mallofré. (2014, April-May). A Model of Calm HCI. In Saskia Bakker, Doris Hausen, Ted Selker, Elise van den Hoven, Andreas Butz, Berry Eggen (Editors) *Peripheral Interaction: Shaping the Research and Design Space*. Workshop at CHI 2014, Toronto, Canada. ISSN: 1862-5207

28. Brown, J. N. A., P. S. Bayerl, Anton Fercher, Gerhard Leitner, Andreu Català Mallofré, and Martin Hitz. (2014, April-May). A Measure of Calm. In Saskia Bakker, Doris Hausen, Ted Selker, Elise van den Hoven, Andreas Butz, Berry Eggen (Editors) *Peripheral Interaction: Shaping the Research and Design Space*. Workshop at CHI 2014, Toronto, Canada. ISSN: 1862-5207

29. Bush, V., "As we may think", Atlantic Monthly, vol. 176, pp. 101-108, 1945.

30. Bystrom, J.U., Hansson, G.-A., Rylander, L., Ohlsson, K., Kallrot, G. and Skerfving, S., 2002, Physical workload on neck and upper limb using two CAD applications. Applied Ergonomics, 33, 63–74.

31. Cadiz, J. J., Venolia, G., Jancke, G., & Gupta, A. (2002, November). Designing and deploying an information awareness interface. In Proceedings of the 2002 ACM conference on Computer supported cooperative work (pp. 314-323). ACM.

32. Carabalona, R., Grossi, F., Tessadri, A., Caracciolo, A., Castiglioni, P., & de Munari, I. (2010). Home smart home: Brain-computer interface control for real smart home environments. Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology, 51.

33. Carabalona, R., Grossi, F., Tessadri, A., Castiglioni, P., Caracciolo, A., & de Munari, I. (2012). Light on! real world evaluation of a P300-based brain–computer interface (BCI) for environment control in a smart home. Ergonomics, 55(5), 552-563.

34. Carroll, L.: The hunting of the Snark, Macmillan, London, (1876)

35. Castelli, G., Rosi, A., & Zambonelli, F.Design and implementation of a socially-enhanced pervasive middleware.

36. Chambers, C. D., and Heinen, K., "TMS and the functional neuroanatomy of attention", Cortex, vol. 22, pp. 114, 2009.

37. Chan, M., Estève, D., Escriba, C., Campo, E.: A review of smart homes - Present state and future challenges, Computer Methods and Programs in Biomedicine 9(I), Elsevier, Amsterdam, pp. 55-81, (2008)

38. Chandak, M. B., Dharaskar, R. (2010). Natural language processing based context sensitive, content specific architecture & its speech based implementation for smart home applications. International Journal of Smart Home, 4(2), 1-9.

39. Chen, C., & Helal, S. (2010). Toward a programming model for safer pervasive spaces. Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2010 7th International Conference on, 52-57.

40. Chikhaoui, B., & Pigot, H. (2010). Towards analytical evaluation of human machine interfaces developed in the context of smart homes. Interacting with Computers, 22(6), 449-464.

41. Choe, B. W., Min, J. K., & Cho, S. B. (2010). Online gesture recognition for user interface on accelerometer built-in mobile phones. Neural Information Processing.Models and Applications, 650-657.

42. Clark, J. D. (1992). African and Asian perspectives on the origins of modern humans. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 337(1280), 201-215.

43. Cohn, G., Morris, D., Patel, S. N. and Tan, D. S. Your noise is my command: Sensing gestures using the body as an antenna. In: Proceedings of the 2011 annual conference on Human factors in computing systems. ACM, pp. 791-800, (2011) 1

44. Cohn, G., Morris, D., Patel, S. N., and Tan, D. S., "Humantenna: Using the body as an antenna for real-time whole-body interaction",in Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, 2012.

45. Cook, D. J. (2012). How smart is your home? Science, 335(6076), 1579-1581.

46. Cook, D. J., & Das, S. K. (2012). Pervasive computing at scale: Transforming the state of the art. Pervasive and Mobile Computing, 8(1), 22-35.

47. Cook, D. J., & Das, S. K. (2007). How smart are our environments? an updated look at the state of the art. Pervasive and Mobile Computing, 3(2), 53-73.

48. Cook, D. J., Youngblood, M., Heierman III, E. O., Gopalratnam, K., Rao, S., Litvin, A., & Khawaja, F. (2003). MavHome: An agent-based smart home. Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on, 521-524.

49. Coomans, M. K. D., & Achten, H. H. (1998, July). Mixed task domain representation in VR-DIS. In Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific (pp. 415-420). IEEE.

50. Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. Nature Reviews Neuroscience, 9(8), 587-600.

51. Csikszentmihalyi, M., & Bennett, S. (1971). An exploratory model of play. American Anthropologist, 73(1), 45-58.

52. Dahl, Y. (2008). Redefining smartness: The smart home as an interactional problem. Intelligent Environments, 2008 IET 4th International Conference on, 1-8.

53. Dainoff, M., & Haber, R. N. (1967). How much help do repeated presentations give to recognition processes? Perception & Psychophysics, 2(4), 131-136.

54. Davidoff, S., Lee, M., Yiu, C., Zimmerman, J., & Dey, A. (2006). Principles of smart home control. UbiComp 2006: Ubiquitous Computing, , 19-34.

55. De Silva, L. C., Morikawa, C., & Petra, I. M. (2012). State of the art of smart homes. Engineering Applications of Artificial Intelligence, 25(7), 1313-1321.

56. Díaz Boladeras, M., Casacuberta Bagó, J., Nuño Bermudez, N., Berbegal Mirabent, J., & Berbegal Mirabent, N. (2011). Evaluación con usuarios finales durante el desarrollo de dos sistemas interactivos orientados a personas mayores.

57. Dimopulos, T., Albayrak, S., Engelbrecht, K., Lehmann, G., & Moller, S. (2007). Enhancing the flexibility of a multimodal smart home environment. Fortschritte Der Akustik, 33(2), 639.

58. Dong, Y., Zhang, B., & Dong, K. (2010). An integrated PLC smart home system in pervasive computing. Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2010 7th International Conference on, 288-291.

59. Dressel, J., & Atchley, P. (2008). Cellular phone use while driving: A methodological checklist for investigating dual-task costs. Transportation research part F: traffic psychology and behaviour, 11(5), 347-361.

60. Drucker, J. (2011). Humanities approaches to interface theory. Culture Machine, 12(0)

61. Edwards, W., & Grinter, R. (2001). At home with ubiquitous computing: Seven challenges. Ubicomp 2001: Ubiquitous Computing, 256-272.

62. English, W. K., Engelbart, D. C., and Berman, M. L., “Display-selection techniques for text manipulation”, Human Factors in Electronics, IEEE Transactions on, pp. 5-15, 1967.

63. Ennis, C., McDonnell, R., & O'Sullivan, C. (2010). Seeing is believing: Body motion dominates in multisensory conversations. ACM Transactions on Graphics (TOG), 29(4), 91.

64. Escobar, A., Hess, D., Licha, I., Sibley, W., Strathern, M., & Sutz, J. (1994). Welcome to Cyberia: Notes on the Anthropology of Cyberculture [and comments and reply]. Current anthropology, 211-231.

65. Evensen, P., & Meling, H. (2009). SenseWrap: A service oriented middleware with sensor virtualization and self-configuration. Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference on, 261-266.

66. Fehérvári, I & Brown, J. N. A. (2012) With a Wave of My Wand: Mobile Gaming with Speech and Orientation-Independent Gestures. Unpublished technical report.

67. Felfernig, A., Mandl, M., Tiihonen, J., Schubert, M., & Leitner, G. (2010). Personalized user interfaces for product configuration. Proceedings of the 15th International Conference on Intelligent User Interfaces, 317-320.

68. Fenton, N. E., & Pfleeger, S. L. (1998). Software metrics: a rigorous and practical approach. PWS Publishing Co. Boston, MA.

69. Ferscha, A., & Resmerita, S. (2007). Gestural interaction in the pervasive computing landscape. E & i Elektrotechnik Und Informationstechnik, 124(1), 17-25.

70. Fiaidhi, J. (2011). Towards developing installable e-learning objects utilizing the emerging technologies in calm computing and ubiquitous learning. International Journal of u-and e-Service, Science and Technology, 4(1)

71. Fitts, P. M. (1947). Psychological research on equipment designs in the AAF. American Psychologist, 2(3), 93.

72. Fitts, P. M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol., 47, 381.

73. Fleury, A., Noury, N., Vacher, M., Glasson, H., & Seri, J. F. (2008). Sound and speech detection and classification in a health smart home. Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, 4644-4647.

74. Fleury, A., Vacher, M., & Noury, N. (2010). SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. Information Technology in Biomedicine, IEEE Transactions on, 14(2), 274-283.

75. Fujinami, K. (2010). Interaction design issues in smart home environments. Future Information Technology (FutureTech), 2010 5th International Conference on, 1-8.

76. Gaudron, J. P., & Vignoli, E. (2002). Assessing computer anxiety with the interaction model of anxiety: development and validation of the computer anxiety trait subscale. Computers in Human Behavior, 18(3), 315-325.

77. Geller, T. (2008). Overcoming the uncanny valley. IEEE Computer Graphics and Applications, 28(4), 11-17.

78. Gilbreth, F. B. (1912). Primer of scientific management. D. Van Nostrand Company.

79. Gordon, J. B., Passonneau, R. J., & Epstein, S. L. (2011). Helping agents help their users despite imperfect speech recognition. AAAI Symposium Help Me Help You: Bridging the Gaps in Human-Agent Collaboration,

80. Grandhi, S. A., Joue, G., & Mittelberg, I. (2011). Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, 821-824.

81. Gu, T., Pung, H. K., & Zhang, D. Q. (2005). A service-oriented middleware for building context-aware services. Journal of Network and Computer Applications, 28(1), 1-18.

82. Hamill, M., Young, V., Boger, J., and Mihailidis, A., “Development of an automated speech recognition interface for personal emergency response systems”, Journal of Neuroengineering and Rehabilitation, vol. 6, pp. 26, 2009.

83. Hausen, D. (2012). Peripheral Interaction: Facilitating Interaction with Secondary Tasks, Proceedings of (TEI 2012), Kingston, 387-388.

84. Hedge, A., Morimoto, S. and McCrobie, D., 1999, Effects of keyboard tray geometry on upper body posture and comfort. Ergonomics, 42, 1333–1349.

85. Heinssen, R. K., Glass, C. R., and Knight, L. A., "Assessing computer anxiety: Development and validation of the computer anxiety rating scale," Comput. Hum. Behav., vol. 3, pp. 49-59, 1987.

86. Helmi, N., & Helmi, M. (2009). Applying a neuro-fuzzy classifier for gesture-based control using a single wrist-mounted accelerometer. Computational Intelligence in Robotics and Automation (CIRA), 2009 IEEE International Symposium on, 216-221.

87. Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. Computers in Human Behavior, 26(6), 1278-1288.

88. Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. Interacting with Computers, 18(2), 227-245.

89. Horrey, W. J., Lesch, M. F., & Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. Accident Analysis & Prevention, 40(2), 675-682.

90. Hosseini-Khayat, A., Seyed, T., Burns, C., & Maurer, F. (2011). Low-fidelity prototyping of gesture-based applications. Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, 289-294.

91. Hurtienne, J., Stößel, C., Sturm, C., Maus, A., Rötting, M., Langdon, P., & Clarkson, J. (2010). Physical gestures for abstract concepts: Inclusive design with primary metaphors. Interacting with Computers, 22(6), 475-484.

92. Ishii, H. (2008). Tangible bits: Beyond pixels. Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, xv-xxv.

93. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, New York, pp. 234-241, (1997)

94. Joselli, M., Clua, E.: gRmobile: A framework for touch and accelerometer gesture recognition for mobile games. Proceedings of Brazilian Symposium on Games and Digital Entertainment, Rio de Janeiro, pp 141-150, (2009)

95. Kaschmitter, J. L. , Shaeffer, D. L., Colella, N. J., McKnett, C. L., Coakley, P. G.: Operation of commercial R3000 processors in the Low Earth Orbit (LEO) space environment, Nuclear Science, IEEE Transactions on , 38(6), pp.1415-1420, (1991)

96. Kim, H. S. (2011, May). Gesture definition approaches and limitations. In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems. ACM

97. Kirk, D., & Stanton Fraser, D. (2006). Comparing remote gesture technologies for supporting collaborative physical tasks. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1191-1200.

98. Kirk, D., Rodden, T., & Fraser, D. S. (2007). Turn it this way: Grounding collaborative action with remote gestures. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1039-1048.

99. Koskela, T., Väänänen-Vainio-Mattila, K., & Lehti, L. (2004). Home is where your phone is: Usability evaluation of mobile phone UI for a smart home. Mobile Human-Computer Interaction–MobileHCI 2004, , 74-85.

100. Kramer, D.R., Halpern, C. H., Connolly, P. J., Jaggi, J. L., and Baltuch, G. H., "Error Reduction with Routine Checklist Use during Deep Brain Stimulation Surgery," Stereotact. Funct. Neurosurg., vol. 90, pp. 255-259, 2012.

101. Kratz, S., & Ballagas, R. (2007). Gesture recognition using motion estimation on mobile phones. Proc. of 3rd International Workshop on Pervasive Mobile Interaction Devices (PERMID'07),

102. Kratz, S., & Rohs, M. (2011). Protractor3d: A closed-form solution to rotation-invariant 3d gestures. Proceedings of the 16th International Conference on Intelligent User Interfaces, 371-374.

103. Kratz, S., Rohs, M.: A $3 gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors. In: Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI'10, ACM, New York, pp. 341-344, (2010)

104. Krishna, Y. B., & Nagendram, S. ZIGBEE based voice control system for smart home. International Journal of Computer Technology & Applications, 3(1).

105. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S.: I'm home: Defining and evaluating a gesture set for smart home control, International Journal of Human-Computer Studies, 69(11), Elsevier, Amsterdam, pp. 693-704, (2011)

106. Kühnel, C., Westermann, T., Weiss, B., Möller, S.: Evaluating multimodal systems: a comparison of established questionnaires and interaction parameters. In: Proceedings of NordiCHI'10, ACM, New York, pp. 286-294, (2010)

107. Kwon, D. Y., and Gross, M., "A framework for 3D spatial gesture design and modeling using a wearable input device", in Wearable Computers, 2007 11th IEEE International Symposium on, 2007, pp. 23-26.

108. Langdon, P., Persad, U., & John Clarkson, P. (2010). Developing a model of cognitive interaction for analytical inclusive design evaluation. Interacting with Computers, 22(6), 510-529.

109. Lecouteux, B., Vacher, M., & Portet, F. (2011). Distant speech recognition in a smart home: Comparison of several multisource ASRs in realistic conditions.

110. Lee, S., & Koubek, R. J. (2010). Understanding user preferences based on usability and aesthetics before and after actual use. Interacting with Computers, 22(6), 530-543.

111. Leitner, G., and Fercher, A., "AAL 4 ALLA Matter of User Experience", Aging Friendly Technology for Health and Independence, pp. 195-202, 2010.

112. Leitner, G., Fercher, A., Felfernig, A., and Hitz, M., "Reducing the entry threshold of AAL systems: Preliminary results from casa vecchia", in COMPUTERS HELPING PEOPLE WITH SPECIAL NEEDSLecture Notes in Computer Science, 2012, Volume 7382, 1st ed., K. Miesenberger, A. Karshmer, P. Penaz and W. Zagler, Eds. Heidelberg: Springer, 2012, pp. 709-715.

113. Leitner, G., Fercher, A.J.: Potenziale und Herausforderungen von AAL im ländlichen Raum. In: Proc. Of Ambient Assisted Living 2011, Berlin, Germany, (2011)

114. Leonardo (1448) Manuscript B, Folios 16r and 37v in the collection of codices of the Institut de France.

115. Lepre, C. J., Roche, H., Kent, D. V., Harmand, S., Quinn, R. L., Brugal, J. P., Texier, P. J. Lenoble, A., and Feibel, C. S. (2011). An earlier origin for the Acheulian. Nature, 477(7362), 82-85.

116. Lesch, M. F., & Hancock, P. A. (2004). Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements?. Accident Analysis & Prevention, 36(3), 471-480.

117. MacKenzie, I. S. (1995). Input devices and interaction techniques for advanced computing. In W. Barfield, & T. A. Furness III (Eds.), Virtual environments and advanced interface design, pp. 437-470. Oxford, UK: Oxford University Press.

118. Makonin, S., Bartram, L., and Popowich, F., "Redefining the" Smart" in Smart Home: Case Studies of Ambient Intelligence," 2012.

119. McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The prevalence of, and factors associated with, serious crashes involving a distracting activity. Accident Analysis & Prevention, 39(3), 475-482.

120. McNeill, D.: Hand and mind: What gestures reveal about thought. University of Chicago Press, Chicago, (1992)

121. Mori, M. (1970). The uncanny valley. Energy, 7(4), 33-35.

122. Myers, B. A. (1998). A brief history of human-computer interaction technology. interactions, 5(2), 44-54.

123. Nakamura, J., and Csikszentmihalyi, M. (2002). The concept of flow. Handbook of positive psychology, 89-105.

124. Nass, C., and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of social issues, 56(1), 81-103.

125. Norman, D. A. (1984). Stages and levels in human-machine interaction. International Journal of Man-Machine Studies, 21(4), 365-375.

126. Olaru, A., Florea, A. M., and El Fallah Seghrouchni, A., "A Context-Aware Multi-Agent System as a Middleware for Ambient Intelligence," Mobile Networks and Applications, pp. 1-15, 2012.

127. Oulasvirta, A, et al. (2007). Communication failures in the speech-based control of smart home systems. 3rd IET International Conference on Intelligent Environments (IE 07), 135-143.

128. Paoli, P., and Litske, H. (1992). First European survey on the work environment 1991-1992. Dublin: European Foundation for the Improvement of Living and Working Conditions.

129. Peterson, J., "Calm technology: Design guidelines," in Umea's 13th Student Conference in Computer Science, pp. 111.

130. Pfajfenberger, B. (1992). Social anthropology of technology. Annual Review of Anthropology, 21, 491-516.

131. Ping, A., et al. (2009). Designing an Emotional Majordomo in Smart Home Healthcare. 2009 International Asia Symposium on Intelligent Interaction and Affective Computing, 45-47.

132. Polanyi, M., Personal Knowledge: Towards a Post-Critical Philosophy. Psychology Press, 1962.

133. Popper, K. R., Objective Knowledge: An Evolutionary Approach. Clarendon Press Oxford, 1972.

134. Raso, I., Hervas, R., Bravo, J.: m-Physio: Personalized Accelerometer-based Physical Rehabilitation Platform. In: Proceedings of UBICOMM'10, Florence, pp. 416-421, (2010)

135. Rauterberg, M. (1996, December). Quantitative Test Metrics to Measure the Quality of User Interfaces. In Proc. of 4th European Conf. on Software Testing Analysis & Review EuroSTAR96, Amsterdam.

136. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress from a Neurobiological Perspective. Business & Information Systems Engineering, 1-9, (2012)

137. Roche, H., Delagnes, A., Brugal, J. P., Feibel, C., Kibunjia, M., Mourre, V., & Texier, P. J. (1999). Early hominid stone tool production and technical skill 2.34 Myr ago in West Turkana, Kenya. Nature, 399(6731), 57-60.

138. Rogers, Y. (2006). Moving on from weiser's vision of calm computing: Engaging ubicomp experiences. In UbiComp 2006: Ubiquitous Computing (pp. 404-421). Springer Berlin Heidelberg.

139. Rouillard, J., Tarby, J.-C.: How to communicate smartly with your house? Int. J. Ad Hoc and Ubiquitous Computing, 7(3), 155-162, (2011)

140. Santangelo, V., Fagioli, S. and Macaluso, E., "The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space," Neuroimage, vol. 49, pp. 2717-2727, 2010.

141. Serina, E.R., Tal, R. and Rempel, D., 1999, Wrist and forearm postures and motions during typing. Ergonomics, 42, 938–951.

142. Sousa Santos, B., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., Madeira, J.: Head Mounted Display versus desktop for 3D Navigation in Virtual Reality: A User Study. Multimedia Tools and Applications, Springer, New York, Vol. 41, pp.161-181, (2008)

143. Stavropoulos, T. G. Vrakas, D. Vlachava, D. and Bassiliades, N., "BOnSAI: A smart building ontology for ambient intelligence," in Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, 2012, pp. 30.

144. Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1567), 1050-1059.

145. Streitz, N., Nixon, P.: The disappearing computer. Communications of the ACM 48(3), pp. 32-35, (2005)

146. Tandler, P. (2001, January). Software infrastructure for ubiquitous computing environments: Supporting synchronous collaboration with heterogeneous devices. In Ubicomp 2001: Ubiquitous Computing (pp. 96-115). Springer Berlin Heidelberg.

147. Taylor, F. W. (1911). The principles of scientific management. New York, 202.

148. Vacher, M., Istrate, D., Portet, F., Joubert, T., Chevalier, T., Smidtas, S., Meillon, B., Lecouteux, B., Sehili, M., Chahuara, P., Méniard, S.: The sweet-home project: Audio technology in smart homes to improve well-being and reliance. In: 33rd Annual International IEEE EMBS Conference, Boston, Massachusetts, USA, (2011)

149. Van Dantzich, M., Robbins, D., Horvitz, E., & Czerwinski, M. (2002, May). Scope: Providing awareness of multiple notifications at a glance. In Proceedings of the Working Conference on Advanced Visual Interfaces (pp. 267-281). ACM.

150. Venkatesh, A. (1996). Computers and other interactive technologies for the home. Communications of the ACM, 39(12), 47-54.

151. Virolainen, A., Puikkonen, A., Kärkkäinen, T., and Häkkilä, J., "Cool interaction with calm technologies: Experimenting with ice as a multitouch surface," in ACM International Conference on Interactive Tabletops and Surfaces, 2010, pp. 15-18.

152. Weiser, M., & Brown, J. S. (1996). Designing calm technology. PowerGrid Journal, 1(1), 75-85.

153. Weiser, M., Brown, J. S.: The Coming Age of Calm Technology, In: Denning, P.J., Metcalfe, R.M. (eds.), Beyond Calculation: The Next Fifty Years of Computing, pp. 75-85, Copernicus, New York, (1997)

154. Weiser, M.: Some Computer Science Issues in Ubiquitous Computing, Communications of the ACM, 36 (7), ACM, New York, pp. 75-84, (1993)

155. Weiser. M.: The Computer for the Twenty-First Century, Scientific American, 265(3), Macmillan, New York, pp. 94-104, (1991)

156. Weiser, M. (1994). The world is not a desktop. interactions, 1(1), 7-8.

157. Weiser, M. (1993). Hot topics-ubiquitous computing. Computer, 26(10), 71-72.

158. Woods, W. A., Bates, M. A., Bruce, B. C., Colarusso, J. J., & Cook, C. C. (1974). Natural Communication with Computers. Volume I. Speech Understanding Research at BBN (No. BBN-2976). BOLT BERANEK AND NEWMAN INC CAMBRIDGE MASS.

159. Zayas-Cabán, T. (2002). Introducing information technology into the home: conducting a home assessment. In Proceedings of the AMIA Symposium (p. 924). American Medical Informatics Association.

160. Zeigarnik, B. (1927). On the Retention of Completed and Uncompleted Activities. Psychologische Forschung, 9, 1-85.