Title

Pervasive Well-being Technology

Author

Paredes Castro, Pablo Enrique

Publication Date

2015

Peer reviewed|Thesis/dissertation

Pervasive Well-being Technology

By

Pablo Enrique Paredes Castro

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

In

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John F. Canny, Chair
Professor Björn Hartmann
Professor Greg Niemeyer
Professor Mary Czerwinski

Fall 2015

Pervasive Well-being Technology

by

Pablo Enrique Paredes Castro

Abstract

Pervasive Well-being Technology

by

Pablo Enrique Paredes Castro

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor John F. Canny, Chair

Well-being is the characteristic of humans to feel well with themselves and their environment. A key driver of wellbeing is to have a healthy mind. Stress management, positive emotions, and empathic social interaction get us closer to this goal. The only sustainable way towards wellbeing is by maintaining a healthy lifestyle. In summary, technologies supporting this objective should measure and intervene "life" itself!

The good news is that we are in a unique position to challenge the way technology can become a driver of wellbeing. Internet of things (IoT), Big Data, Affective Computing and Critical Engineering, are the key enablers. With pervasive data, interventions can evolve from being reactive to predictive. We could design long-term interventions that foster good habits, resilience, and personal growth. We encapsulate technologies in two parts: "Conceptualization and Sensor-less Sensing" and "Opportunistic Interventions." The former uses data streams to investigate and measure human traits and trends. The latter repurposes popular apps, devices, and environments into effective interventions.

In summary, we aim at designing well-being technology that could survive the chasm of adoption and attrition.

Dedicated with love to my wife Claudia (Clau), my babies Manuela (Manu) and Diego Pablo, my parents Enrique and Cecilia, my brother Juan Carlos, my sister Maria Cecilia. You made this possible. Thanks forever!

Contents ii

List of Figures vi

List of Tables x

Introduction 1

Acknowledgements 3

Part 1: Conceptualization and Sensing: 5

1. Qualitative Insight Mining 7

1.1 INTRODUCTION ... 8
1.2 RELATED WORK ... 10
1.2.1 Keyword Search ... 10
1.2.2 Semantic Similarity ... 10
1.2.3 Qualitative Data Analysis Tools ... 11
1.3 METHODOLOGY ... 11
1.3.1 LiveJournal Corpus ... 11
1.3.2 Data Preprocessing ... 11
1.3.3 word2vec ... 12
1.3.4 Querying ... 13
1.4 ITERATIVE CO-DESIGN PROCESS ... 15
1.4.1 Initial Concept Formation ... 15
1.4.2 Early Idea Validation ... 17
1.4.3 Low Fidelity (Lo-Fi) Prototype ... 17
1.4.4 Testing on more Data ... 22
1.4.5 Mid Fidelity (Mi-Fi) Prototype ... 23
1.5 FUTURE WORK ... 26
1.6 CONCLUSION ... 27

2. Game & Narrative Design for Behavior Change 29

2.1 INTRODUCTION ... 30
2.2 PREVIOUS WORK ... 30
2.3 THE CHALLENGE: EFFICACY + EFFICIENCY ... 31
2.4 DESIGN METHODOLOGY ... 32
2.5 GAME PROTOTYPE EXAMPLES ... 33
2.6 NARRATIVE PROTOTYPE EXAMPLE ... 36
2.6.1 Mental Health Background ... 36
2.6.2 Related Work ... 39
2.6.3 Prototype Movie: “A Rainy Day” ... 39
2.6.4 Uses of Therapeutic Machinima ... 42
2.6.5 Future Research ... 42

2.7 PRINCIPLES FOR CONCEPTUALIZATION 43
2.7.1 Understanding Disempowering Narratives 43
2.7.2 Externalizing Problems 43
2.7.3 Game Dynamics as Interventions 45
2.7.4 Narratives as Interventions 45
2.8 CONCLUSION 45

3 Sensor-less Sensing 46
3.1 INTRODUCTION 47
3.2 BACKGROUND 47
3.2.1 Arousal and Stress 47
3.2.2 Physiological Affect Sensing 48
3.3 AROUSAL FROM COMPUTER PERIPHERALS 49
3.3.1 Mouse Movement 49
3.3.2 Mouse Pressure 52
3.3.3 Keyboard 53
3.4 SENSING AFFECT WITH TEXT MINING 57
3.5 MOVING FROM SENSING TO INTERVENTIONS 57
3.5.1 Usability Studies 57
3.5.2 Stress and Emotion from Sensor and Contextual Data 58
3.5.3 Integrated Model and Data Acquisition 58
3.5.4 Causal Inference 58
3.6 ADAPTIVE INTERACTIONS AND INTERVENTIONS 59
3.6.1 Feedback and Interface Adaptation 59
3.6.2 Emotional Regulation and Psycho-education Interventions . 59
3.6.3 Foreground and Background Interventions 60
3.7 CONCLUSION 60

Part 2: Opportunistic Interventions 61

4 Repurposing mobile and wearable systems 62
4.1 INTRODUCTION 63
4.2 BACKGROUND AND RELATED WORK 63
4.3 IMPLEMENTATION 63
4.3.1 Prototypes 63
4.3.2 Common Use Scenario 66
4.4 STUDY DESIGN 66
4.4.1 Objective Data 66
4.4.2 Subjective Data 67
4.5 EVALUATION 67
4.5.1 Baseline 67
4.5.2 Randomized Experiment 68
4.6 RESULTS 68

4.7 IMPLICATIONS FOR DESIGN 71
4.7.1 Suite of Interventions and Context 71
4.7.2 Design Improvements 72
4.7.3 Long-term (Real-Life) Usability 73
4.8 CONCLUSION 73

5 Harvesting web interventions **74**
5.1 INTRODUCTION 75
5.2 BACKGROUND WORK 76
5.2.1 Cognitive Behavioral Therapy (CBT) based Technology 76
5.2.2 Stress Technology Intervention R&D 76
5.3 STRESS MANAGEMENT SYSTEM DESIGN 77
5.3.1 Micro-Intervention Authoring System 77
5.3.2 Intervention Recommender System 80
5.3.3 Mobile App Interface 83
5.4 STUDY DESIGN 85
5.4.1 Phones and Screening Procedures 85
5.4.2 Experiment Design 86
5.4.3 Gratuity and Incentive Policies 86
5.5 RESULTS 86
5.5.1 Descriptive Data 87
5.5.2 Qualitative Results 91
5.5.3 Quantitative Results 92
5.6 IMPLICATIONS FOR DESIGN 94
5.6.1 Intervention Design 94
5.6.2 Intervention Recommendation 95
5.6.3 Future Research 96
5.7 CONCLUSION 96

6 Multi-modal Interventions **97**
6.1 INTRODUCTION 98
6.2 BACKGROUND 99
6.2.1 Circumplex Model of Affect 99
6.2.2 Auditory Affective Expressivity 100
6.2.3 Haptic Affective Expressivity 100
6.2.4 Multisensory Affective Response 100
6.3 PRIOR WORK 101
6.3.1 Audio and Haptic Research 101
6.3.2 Vibrotactile Interfaces 101
6.4 EXPERIMENT DESIGN 102
6.4.1 Vibrotactile Wearable Hardware 102
6.4.2 Stimuli Generation 102
6.4.3 Conditions 104

6.5 RESULTS 104
6.5.1 Quantitative Analysis 104
6.5.2 Qualitative Analysis 109
6.5.3 Usability Assessment 111
6.6 IMPLICATIONS FOR DESIGN 113
6.7 FUTURE WORK 114
6.8 CONCLUSION 115

7 Urban Ambien Interventions – Fiat Lux 116
7.1 INTRODUCTION 117
7.2 PREVIOUS WORK 118
7.2.1 Urban Interactive Light Design 118
7.2.2 Light Interaction and Expressivity 118
7.2.3 Implicit Interaction and Approachability 118
7.3 LIGHTING SYSTEM DESIGN 119
7.4 METHODOLOGY 119
7.4.1 Location 119
7.4.2 Method 121
7.5 EXPERIMENT 1: INTERACTIVE LIGHT FACTORS 123
7.5.1 Experiment Design 123
7.5.2 Quantitative Analysis 124
7.5.3 Post-test Qualitative Analysis 126
7.5.4 Card Sorting Analysis 127
7.6 EXPERIMENT 2: INTERACTIVE LIGHTS VERSUS ALWAYS ON 132
7.6.1 Experiment Design 132
7.6.2 Quantitative Analysis 132
7.6.3 Post-test Qualitative Analysis 133
7.6.4 Card Sorting Analysis 134
7.7 EXPERIMENT 3: BEFORE + SLOW PRIMITIVES 136
7.7.1 Experiment Design 136
7.7.2 Quantitative Analysis 137
7.7.3 Post-test Qualitative Analysis 139
7.7.4 Card Sorting Analysis 139
7.8 IMPLICATIONS FOR DESIGN 140
7.9 CONCLUSION 144

8 Conclusion 145
8.1 SUMMARY........ 145
8.1.1 Part 1 – Conceptualization and Sensing 145
8.1.2 Part 2 – Opportunistic Interventions 145
8.2 INTERVENTION RESEARCH DRIVERS 146
8.3 VISION 146

References 148

List of Figures

1.1 Example based on a real query, on how the queries evolve in response to the returned answers. The flow is somewhat analogous to semi-structured interviews, where interviewers must modify their line of questioning on the fly to respond to answers received. The example is linear for simplicity of presentation, but the interviewer can easily follow a line of questioning and then return to an earlier query as many time as desired. 16

1.2 Example of the effect of filters in the query results on the full data. Top right displays the top 8 results for the query "family holiday". To see more in depth responses, we can filter out the sentences with less than 30 words, see results in the bottom left. Since we perform a semantic query, we may choose to exclude the words "family" or "holiday" from the results. In the top right panel, we exclude the word "family". Finally, we can define a weighting scheme on the query words. In the bottom right panel, we increase the weight on the word "family" to obtain results that relate more to "family" than "holiday". Note that each result belongs to a full post, which we retrieve to the interviewer, e.g., the original post of the highlighted result is shown in Figure 1.3. 20

1.3 Public post on LiveJournal.com, corresponding to the highlighted result in Figure 1.2. By viewing the full text in its original context, the interviewer is able to explore a response in greater depth. 21

1.4 Example queries and results ran on 100% of the data, and with distinct filters. 25

2.1 Efficacy + Engagement game design dualities. 32

2.2 Monsters Game Screen. 34

2.3 Scheherazade's Word Game Screens. 35

2.4 Semester adventure game screen. 36

2.5 Original CBT cartoon versus machinima video screenshot. 38

2.6 Excerpts from original CBT manual and machinima movie script. 40

3.1 Mouse displacement damping ratio for 20 different motion tasks for mental Non-stress and Stress conditions [22]. 51

3.2 Affective and Stress Measurements using 7-point Likert scales. 52

3.3 Individual differences between the averaged differential (stress – calm) signal per participant. *The difference was computed from significantly different distributions (W, p<0.05). …………………………………………… 55

4.1 An SMS message used to prompt the closest contacts in the user's (inner) social network. ……………………………………………………………….. 64

4.2 One of the games we used involved maneuvering a ball in a maze. ………. 64

4.3 Vibrotactile bracelet with two eccentric rotating mass (ERM) motors for guided breathing and acupressure stimulation. ……………………………… 65

4.4 Accupressure points for relaxation. …………………………………………… 66

4.5 Experiment Stages. …………………………………………………………….. 68

4.6 Subjective stress levels for each intervention across the different stages on part 2. ……………………………………………………………………... 69

4.7 Comparison of the four interventions. Social Networking and Guided Breathing had the most uniform distribution of usability traits. ……………… 69

4.8 Electro-Dermal Activity (EDA) oscillating around 540 mV. ………………….. 70

4.9 Biometric signals comparison between Calm and Stress stages. ………….. 71

5.1 Mobile app sample screens: a) subjective stress assessment; b) intervention title, icon, slogan and prompt. …………………………………… 84

5.2 Experience Sampling Method (ESM): a) Self-report message to be clicked by user; b) Circumplex quadrant selection; c) Self-rating completion with option to launch micro-intervention if desired. ………………………………. 85

5.3 Machine Learning selected interventions. …………………………………… 87

5.4 Distribution of interventions per participant. …………………………………. 87

5.5 Fraction of interventions for which users selected the recommended intervention. ………………………………………………………………….. 88

5.6 Stress deltas (stress before – stress after). ………………………………… 89

5.7 Daily users per day per experiment condition. ……………………………… 89

5.8 "Ideal" users (stress self-rating before intervention > 0.5 and intervention duration > 60 sec) a) Stress after versus stress before interventions; b) stress delta versus intervention duration. 90

5.9 Subjective ratings (likeability and efficacy). ………………………………… 91

5.10 Depression (PHQ-9) quantitative results. ………………………………… 93

5.11 Coping Strategies Questionnaire (CSQ). 94

6.1 Circumplex Model of Affect by Russell [16]. 99

6.2 Basic haptic effects with two vibrating motors as described in [12]. 101

6.3 a) Prototype wristband made of Velcro straps covering the vibrating motors and attached to Arduino Uno. b) Motor placement (1.1 inches apart). 102

6.4 Stimuli affect map: a) Haptic, b) Sound. (o) represents average points for all 40 users; (x) represents the centroids; colors represent the different stimuli: black = happy, red = anger, blue = sadness, green = relaxed. 105

6.5 Multiple comparisons (with Bonferroni's correction). Sound mode's valence is significantly different than Haptic mode or Combined (Haptic + Sound) mode. 107

6.6 Multiple comparisons (with Bonferroni's correction). Happy haptic (High Valence + High Arousal) combined with Anger sound (Low Valence + High Arousal) was significantly different. 108

6.7 Multiple comparisons (with Bonferroni's correction). Only Anger and Relaxed present significantly different valences. 109

6.8 Usability metrics for Sound, Haptic and Combination (Sound + Haptic) conditions. 112

7.1 Experimental setup a) metrics: light radius = 1.12m, distance between lights = 2.7m, hallway width = 4m, hallway length = 30ft, base luminance with lights off = 2lx. Note: low illumination not displayed in this image; b) User walking with all lights on. Max luminance per light = 40 lx. Note: Image captured with high sensitivity camera, appears to be brighter than reality.... 120

7.2 Hard Lighting Interactions. 122

7.3 Circumplex Model of Affect (CMA) mapping of the Arousal and Valence metrics for each condition in study number 1. 122

7.4 Energy consumption (Wh) in red, contrasted with Valence (baseline corrected) in blue for each light condition. 125

7.5 Board used to sort the different cards with places into the best affective classification. 128

7.6 Example of the light conditions board. Used to match places with light conditions. 129

7.7 Interaction between Valence (feelings) about a place and (a) Timing and

(b) Ramp factors. …………………………………………………………………… 131

7.8 Energy consumption (Wh) –red- contrasted with Valence (baseline corrected) –blue- for each light condition. ……………………………………… 133

7.9 Interaction between Valence (feelings) about a place and (a) Timing and (b) Ramp factors. …………………………………………………………………….. 135

7.10 Interaction between Valence (feelings) about a place and (a) Timing and (b) Ramp factors. …………………………………………………………….. 137

7.11 Interaction between Valence (feelings) about a place and the number of Look Ahead Lights. …………………………………………………………….. 140

7.12 Inverse U curve about a place and (a) Timing and (b) Ramp factors. …. 141

List of Tables

2.1 Behavior Change and Narrative topics taught in addition to Game Development and Human Centered Design topics. ………………….... 33

2.2 Movie script elements of “A Rainy Day”. ……………………………….. 41

5.1 Micro-interventions design matrix. Therapy groups are subdivided into individual and social groups. Each therapy group has a friendly icon and name. The micro-intervention format consists of a Prompt plus a URL. ………………………………………………………………… 79

5.2 User data and their parameters. ………………………………………… 81

5.3 Sensory features collected on the phone. ……………………………… 81

5.4 Distribution of participants and interventions for the different conditions of the study. …………………………………………………… 86

5.5 Reported learning. ………………………………………………………… 92

6.1 Top four haptic patterns used for the expression of emotions[9]. …… 100

6.2 Circumplex Model of Affect with haptic and sound mappings per quadrant. ……………………………………………………………………… 103

6.3 Experiment conditions. …………………………………………………… 104

6.4 Three-way ANOVA for Arousal values. ………………………………… 106

6.5 Three-way ANOVA for Valence values. ………………………………… 106

6.6 Preferred communication modalities. …………………………………… 113

7.1 Counts of places with respect to the feelings and level of excitement. 130

7.2 Counts of Places with respect to the feelings and level of excitement. 134

7.3 Counts of places with respect to the feelings and level of excitement. 139

7.4 Design implications for optimizing affect versus energy consumption in urban interaction illumination systems. ………………………………. 142

Introduction

A vision shared by all the fields of computer science is to improve the living conditions for human beings. Computers have affected areas such as labor conditions, access to information, entertainment and science. Computers have touched almost every aspect of human existence. They have brought notable improvement to our living standards. There are frontiers that we need to conquer, and computers should play a significant role in that regard. Mental health is one of those. Computers are not yet able to know how we feel, cannot answer what we think, and do not explain our behaviors.

Despite their pervasiveness, computers have not driven large-scale adoption of well-being technology. In fact, the adoption of mental health continues to be so low nowadays that we should count this as an open problem. 30% of the population will need mental health treatment at some point in their lives. However, only, 10% will see a mental health specialist. Out of those only 30% will get treated. From those, only 10% will complete treatment and sadly, 70% of them will relapse. 0.07% is a staggering low efficiency of such a fundamental need. Our approach in this work is not to automate mental health theory. We want to use Human-Computer Interaction (HCI) theory and research to propose novel tools. We want to improve intervention adoption while maintaining or improving their efficacy. We can achieve this by repurposing popular technologies such as web apps, LEDs, and wearable devices (a.k.a. wearables).

Furthermore, the need for intervention research cannot be in tandem with sensing research. The drivers of adoption, engagement and compliance are fundamental problems of human-centered design. For example, novelty is a driver of adoption. However, it is a double edge sword. Target populations drawn to an intervention are quick to realize they are not ready to adopt it. The technology reminds them of their suffering, and the incremental effects are not valuable in the short term. For those engaged, once novelty dies, even an efficacious intervention can die. People would abandon even good designs over time. All this shows the need to not only design interventions, but also sensors. We need sensing technologies that are stealth and pervasive. This way we use long-term adoption timelines, or wait for the right time to launch an intervention.

Part 1 discusses the initial stages of intervention design. It proposes methodologies and tools that help conceptualize interventions that are engaging and efficacious. It also discusses a novels stealth approach to sensing, which we coin "Sensor-less Sensing." Part 2 presents different opportunistic well-being interventions. It discusses how popular tech design, popular devices, and popular places drive adoption.

In summary, we present a group of work that draws on HCI theory and creativity. We aim at breaking the mold of automation of clinical psychological treatments. We take advantage of well adopted (popular) interaction designs (apps, games, etc.) to make interventions desirable. We hope to inspire health care to think beyond traditional clinical paradigms. Ultimately, we want to influence the way we design technology to focus it on well-being.

Acknowledgements

First I want to thank my advisor, Professor John F. Canny. I am grateful for picking me up among hundreds, despite coming from a non-traditional career. Through him I learned to do rigorous science. It was not easy, it took effort to discover how hard it was, but how pleasing it was at the end.

Special thanks to my parents-in-law, Regulo and Nancy. Your help to build our little house in San Leandro was invaluable! Thanks to my dear neighbors Tim and Jennifer, you made our stay in California a delight and an inspiration.

I want to thank my academic and life mentors:

Prof. Eric Brewer - You guided my passion with such smart vision. Your incredible knowledge of technology is only surpassed by your human kindness. Thanks for making me a putative son of TIER!

Dr. Margie Morris - without your help during my application phase I would have not had a strong enough application. I hope we can continue to collaborate in years to come.

Dr. Sunny Consolvo – I managed to understand the value of a strong will and keen intelligence. She is a natural leader. Thanks for showing me the power of usability design in a world full of technicians!

Dr. Mary Czerwinski – Thanks for believing in me. I still think we chose the most interesting and rewarding topic. There is so much we need to discover, I just hope we will always be able to come with new projects together.

Prof. Greg Niemeyer - You are one of the most kind, dynamic, creative and interesting person I met at Berkeley. Without your inspiration, patience and faith in me, making it through the last year would have been very hard.

Prof. Bjoern Hartmann - You became my "ad hoc" advisor and my key funding partner support for the latter part of my career. You are without any doubt on of the brightest minds in HCI. I hope we will not stop collaborating

Prof. John Chuang - What a pleasure it was to work with you. You always valued my somewhat eclectic set of abilities. I always admired your ability to bring together so many stake holders to the wearables class. I am so glad we became co-authors. Let's keep collaborating!

Prof. Ricardo Munoz – Thanks for seeing in me an agent for change and for welcoming into your research groups at UCSF. I learned so much through you and your colleagues!

Prof. Adrian Aguilera – Thanks for your constant support and insights in the wonderfully complex world of clinical psychology.

Prof. Stephen Schueller – Thanks for your guidance in the early stages of my thesis and for your inspiring passion for the use of technology in psychology.

I want to thank all my colleagues at the Berkeley Institute of Design (BID), Google, Microsoft Research (MSR) and the Technology and Infrastructure for Emerging Regions (TIER) groups. All of you had been a strong inspiration for me to pursue novel and challenging ideas. I cannot be more lucky to have worked with such a talented interdisciplinary team. Special thanks to Keng-hao, Reza, David and Kristin at BID and Rani and Asta at MSR. You all made me challenge my ideas to the limit.

Thanks to my colleagues in the Social Apps lab, my house for the last year. Eduardo Calle, Marijo Recalde, Soham Shah, Mackenzie Alessi, you all made a very difficult project implementable and so enjoyable! The city of San Leandro will always be in debt with you all.

Javier Rosa, thanks for becoming such an unlikely friend. I think destiny brought us together, and we managed to create a strong friendship based in pure creative tension. Thanks for your intelligence, technical prowess and true passion for technology. I will always count you as one of the highlights of my experience at Berkeley.

Ryuka Ko, thanks so much for so many things. You were not only my co-author, but my colleague and councilor. I just hope to see you playing music in the very best outlets. You are so talented. If music is your destiny, science will be missing a great talent. If science becomes your home, music will be missing one.

Dav Clark, thanks so much for your frank and open friendship. You were, no doubt a deep source of creative trouble for my soul! Your keen observations helped shape a much more critical mind.

David Sun, we worked so hard together. There is so much to be said about how much we learned by making mistakes. Thanks for working with me and getting so interested in the topic that brought me to Berkeley.

Thanks to my co-authors and Berkeley Institute of Design fellow collaborators. I call you all the fantastic five. Ana Rufino Ferreira, Cory Schillaci, Dennis Xing, Pierre Karashchuck and Gene Woo. What a magical experience to have so much talented interested in moving forward a crazy idea around a difficult but relevant problem, mental health. We came a long ways and I truly hope we can make our creation either a product or a fantastic research tool. Let’s keep walking together!

To my co-authors in the Synestouch paper. Ryuka Ko, Arezu Aghaseyedjavadi, Linda Babler, Eduardo Calle, Thanks so much for your passion. We managed to do so much with so little. We managed to publish a paper that despite its complications, helped open the door to understand multi-modal expressions. I am so happy to have met you all.

Matt Chan, my first co-author ever! Thanks for your collaboration. Our CalmMeNow paper opened me so many doors! Thanks for your interest in mental health!

Eduardo Calle, thanks so much for sticking with science ever since we started that ambitious project in Ecuador. I hope to see you in a few years fulfilling your dream to get a PhD in robotics!

Part 1
Conceptualization and Sensing

Designing interventions is more an art than a science. Those involved in mental health, know how hard it is to make an effective theory adoptable. The first part is to understand the underlying phenomena and drawing insights. Traditional research methods are usually expensive. Qualitative approaches generate preliminary and descriptive valuable insights. Qualitative approaches are rarely scalable. We describe a novel system based on semantic searching that scales up qualitative exploration. We used a corpus of data with rich emotional content, livejournal.com. Our preliminary results show a huge potential that can even transcend mental health. Qualitative researchers, in general, could use this tool to gather intervention design insights. Furthermore, this tool could potentially even measure human traits captured in personal free writing.

Insights are not enough to design appropriate interventions. Design methods that leverage existing therapeutic knowledge are necessary. When designing behavior change apps, it is not sufficient to build new technology. Many times, an intervention tries to correct something broken. An understanding of the "negative" scenario and successful therapies will improve the design outcome. We studied projects from student groups focused on this topic. Their outcomes helped establish some design principles for effective and efficient interventions.

Finally, "Sensor-less Sensing" is our proposal to use personal data as sensors. We propose to repurpose signals from peripherals and social media to track mental states. Unobtrusive Sensing is a must do to enable pervasive wellbeing interventions. There are advantages in leveraging daily use computing devices and social media as “sensors”:

> *Accessibility*. Peripherals, such as mice and keyboard, are indispensable to interaction with graphic user interfaces.
>
> *Unobtrusiveness*. There is no need for sensors on the body, especially in semi-structured office spaces.
>
> *Long-term, in-situ monitoring*. People spend considerable amounts of time using computers. We can unobtrusively monitor and provide feedback.
>
> *Application and content-neutral monitoring*. Mice motion is neutral to the application and the content with which the user is interacting. This capability improves adoption and reduces privacy concerns.

Chapter 1 shows a novel way to research qualitatively social media using semantic searching. This approach can help study many daily life topics, including mental health ones. This approach transcends sensing, as it can also help discover potential interventions. Chapter 2 describes a method used for the design of creative behavior change interventions. Chapter 3 introduces the use of the PC peripherals such as mice, keyboards, and microphones as sensors.

Chapter 1
Qualitative Insight Mining

In this chapter we describe a novel tool to open up new windows of exploration for qualitative researchers. Through the use of semantic search algorithms applied to a large corpus of data, obtained from LiveJournal.com, we are able to extract valuable information that qualitative researchers can use in their methodology. We describe the design processes that lead us to the development of this tool. We analyze the results of an open user study with qualitative researchers, which helped us confirm the value of the tool and lead to the development of additional features. Finally, we present comparisons between results obtained by modifying certain input parameters such as: percentage of data used, number of results returned, filtering, and word weighting.

1. INTRODUCTION

Semi-structured interviews are a central tool in user research and the social sciences more generally. They allow for rich and open-ended exploration of user attitudes, preferences, desires, fears and values. From a design perspective, their value relies in the ability to generate insights. While they are of tremendous value, they also have difficulties:

- They are expensive in time and resources. They often involve travel by one or both parties, recording and transcription of content, subject reimbursement, and later effort by the interviewer to form a synthesis from individual inter- views.
- They are real-time which affords continuity in the thoughts of the interviewee, but also challenges the interviewer to decide instantly what topics to pursue.
- They typically involve a relatively small sample of subjects. While one obtains a breadth of responses it is difficult to know how widely-shared these are.
- They are normally time-bounded. Parties make a prior commitment to the duration of the interview, which cannot be adjusted to follow promising threads that emerge late in the interview.

In this chapter, we explore the use of Big Data technologies to provide an complementary tool that serves some of the same goals of semi-structured interviews and other methodologies of discover. Specifically, we explore recently developed deep semantic embedding techniques to discover relevant posts from a large corpus of user-generated content. We use a corpus of public posts from the LiveJournal site. LiveJournal is a social media site where users maintain a personal online journal or diary. LiveJournal posts tend to be longer than other sites, and the journal format encourages externalization of user attitudes and values. Explicit sentiment is attached to some posts via an emoticon system. The site contains many themed areas around topics such as health and lifestyle. Even among the public posts there are many where users discuss significantly life challenges and seek and provide support for others. The content does not cover every possible topic, but the corpus is quite large and rich and coverage of rarer topics grows as the corpus grows. We describe a tool which allows an “interviewer” to explore this corpus via a series of “questions” similar to a semi-structured interview. Each question retrieves a set of relevant sentences from various parts of the corpus, and the containing posts can be retrieved to provide context around them.

Within the LiveJournal corpus are significant tracts of text where users discuss attitudes, values and personal experiences around various topics. We argue that these tracts are similar to the responses users might give in a semi-structured interview. They even occur in the context of that user's history of posts, which can be explored to better understand their background. The challenge then is to separate the relevant tracts from other parts of the corpus. This is where deep semantic technologies come in. Word2vec is a neural semantic embedding method that has been shown to perform well at a variety of semantic tasks. "Semantics" here includes not just propositional content, but sentiment and even style. We argue that use of these technologies provides a much richer form of matching than traditional keyword-based search and even first-generation semantic search techniques such as Latent Semantic Indexing (LSI). We bolster this argument with our experiments which showed that perceived quality of the retrieved results was improved by *filtering out* surface-matching (i.e. keyword- matching) posts from semantically-matching posts.

While this approach is not a replacement for semi-structured interviews, it can serve some of the same goals. This is especially true for topics that are more general and therefore well-covered in the corpus. For these topics, there are potential advantages of this approach over classical interviews, which include:

Economy. Our approach is very economical compared to live interviews. No travel is involved, no transcription, no subject payments.

Interviewer Reflection. Interviewers can take as much time as needed to compose questions. There is no cost to pursuing side themes, or with "dead-end" themes. Interviews can last as long as desired, be paused/restarted etc.

Representativeness/Diversity. The number of users who will post about a popular topic (e.g. health, transportation) is large compared to the number of users who would be involved in face-to-face interviews. By exploring up and down the set of relevant sentences, the interviewer gains a sense of typical attitudes.

Interviewer Control. Our system provides a variety of controls and filters over the retrieved results. With practice, this allows the interviewer to more quickly get the kinds of responses they are looking for, compared to reformulating questions in a live interview.

We present evidence that the value of this approach improves with the size of the corpus and the effectiveness of the semantic embedding method. With the recent rapid progress in scalability and accuracy of semantic embedding methods, this means that our approach will improve in usefulness with these developments.

1.2 RELATED WORK

1.2.1 Keyword search

Keyword searches were the first approach to mining a large corpus of data for relevant information. Exploration techniques using single-word queries date back to keyword-in-context indexing (KWIC) [78, 40] which provides snippets of text surrounding the search query and a key for reading the text in its original context. This model is still relevant, but the results are difficult to parse and recent work has focused on visualization methods such as Word Tree [149] which aggregates results with identical word sequences and displays them in a tree-like structure. Later visual text exploration tools based on the Word Tree visualization paradigm include WordSeer [104] which makes it easier to navigate a corpus by facilitating switching between different views, adding summary statistics and applying filters (such as date ranges or other metadata). Other refinements included imposing specific grammatical requirements on the results [103].

A standard goal of semi-structured interviews is to obtain non-obvious insights, responses pertinent to the original topic but not expected by the interviewer. Most such interview responses are not expected to explicitly contain words from the interviewer's questions. For this purpose, keyword search techniques can be too restrictive. Some efforts have been made to search a corpus directly by document level themes. One such example provides a flexible visualization and exploration tool [35] based on topic modeling, a technique to identify latent themes across a collection of documents [13]. As an alternative to the latent concept approach, the relatedness of texts has also been computed using explicit semantic analysis on natural concepts as defined by humans via Wikipedia [45]. An- other attempt to leverage the Wikipedia corpus for semantic annotation of content is described in [89].

1.2.2 Semantic Similarity

Researchers have also made efforts to developing a semantic similarity metric applicable at the word or sentence level. Schemes for measuring similarity at the word level include LSI and Point-wise Mutual Information and Information Retrieval (PMI-IR) [141], both of which learn semantic relation- ships based on co-occurrence of words in a training corpus. More recently, distributed word embeddings learned with recurrent neural networks have become state of the art through algorithms like word2vec, which we use here and describe in the Methodology section. Building on these efforts, various models to compute short text similarity based on word-level semantic similarity were also developed [90, 68]. A recent pa- per also attempts to develop direct embeddings at the sentence level using long short term memory (LSTM) neural networks, called skip-thought vectors [72].

1.2.3 Qualitative Data Analysis Tools

In recent years, Qualitative Data Analysis (QDA) tools have been quite helpful to streamline qualitative research. Tools such as MaxQDA [86], Atlas.ti [6], and N-Vivo [108] are among the most popular ones. These tools provide a wealth of functions to make coding content faster and more precise. They all help generate insights drawn from different sources of data such as videos, text, annotations and recordings. However, these tools have not shown the possibility to perform semantic searching yet. They also do not have direct knowledge connections to sources such as social media or libraries. We are aware that some of these products have not yet succeeded to provide semantic search. We are confident that our tool could complement these tools as a plug-in.

1.3 METHODOLOGY

1.3.1 LiveJournal Corpus

For this study, we obtained a data set consisting of all public English-language LiveJournal posts as of November 2012 in raw XML format. As previously mentioned, LiveJournal is a social networking service where users can keep a blog, journal, or diary. Users have their own journal pages, which show all of their most recent journal entries. Each journal entry can also be viewed on its own web page which includes comments left by other users.

As of 2012, LiveJournal in the United States received about 170 million page views each month from 10 million unique visitors. Our data contains about one million users with a total of 64,326,865 text posts. The median sentence length is 10 words and, as expected, the distribution of sentence lengths follows a power law. Of the users that provided their date of birth, the majority were in the 17-25 age group. Users who chose to identify their location were primarily in the United States (72%), with significant populations in Russia, Canada, the United Kingdom, and Australia. Additionally, users were able to indicate their binary gender; of those who did so 45% identified as male and 55% as female.

1.3.2 Data preprocessing

We started with a raw XML dataset which consists of more than 9 billion words in over 500 million sentences. In order to extract semantic meaning and run synthetic interviews through the use of natural language processing (NLP) algorithms, we implemented a data pipeline that allows us to extract and clean the post text in a computationally efficient form.

The first step is to tokenize the corpus using FLEX (Fast Lexical Analyzer), a scanner which maps each word in the corpus to a unique number and saves the mapping in a dictionary. For efficiency, we keep only the 994,949 most common words and discard the rest (all of which occur fewer than 41 times in the entire dataset). Care is taken to properly tokenize numbers and emoticons such as “:-)”, “=<” or “>:P”.

After removing non-textual content such as images, hyperlinks, and XML formatting data, a bag-of-words (BOW) representation is saved for each sentence from the posts as a column in a sparse feature matrix. The rows of this matrix correspond to the entries in the dictionary above, and the values in the matrix are the counts of each entry for each sentence. For each sentence, we also store the ID of the post it belongs to and the user who wrote it so that the original post can be found online. Finally, we keep the tokenized sentence data to display results when performing the query.

1.3.3 word2vec

We explore synthetic interviewing by employing word2vec [91, 92], a popular word embedding model which has been successful in a variety of NLP applications, including analogy tasks, sentence completion, machine translation [85], and topic modeling [32]. Word embedding models map each dictionary word into a lower dimensional continuous vector representation. With word2vec, this embedding is learned automatically from a sample corpus using a recurrent neural network.

The true power of the technique is that the resulting feature space demonstrates semantic structure, e.g. vec(“king”) − vec(“man”) + vec(“woman”) has a greater co- sine similarity to vec(“queen”) than to the vector of any other word in the dictionary. We can also generalize this comparison technique beyond the scope of single-words. By performing vector addition on the word vector embeddings of corresponding words and normalizing the resulting sum, we can compare the semantic proximity of complete sentences. This procedure is described in detail below.

We use two embedding models trained on two different corpus. First, we train our own embedding model on the corpus of LiveJournal posts using the Skip-gram with negative sampling (SGNS) implementation of word2vec as recommended in [92]. This model gives 300-dimensional embeddings for the 994,949 words in our dictionary. Conveniently, Mikolov et. al published a pre-trained SGNS word2vec model for open source access [93]. This model was trained on a 100 billion-word Google News corpus, and gives 300-dimensional embeddings for 3 million unique words. All of the words from our LiveJournal dictionary that are not included in this model are mapped to zero vectors. Common stop words (i.e. “the”, “an”, “who”) provide little semantic information and are also mapped to zero vectors.

1.3.4 Querying

The goal is to perform semantic searches on the dataset from queries which can range from individual words to full sentences. To do this, we find the similarity of sentences based on the embeddings of their individual words. Specifically, we define a sentence embedding to be the mean of its constituent word vector embeddings. The interviewer's question is also converted to a sentence embedding and tested on the data set using cosine similarity. Note that multi-sentence queries are also possible by treating the entire query in the same way. Word embeddings have been employed in more recent and sophisticated approaches, which outperform our similarity method when used in short text strings [68]. However, our approach has sufficient power for our goal while favoring rapid retrieval of the best matches from a large corpus.

In order to query the data, we multiply the word2vec embed- ding matrix and the sparse bag of words matrix described above to get a semantic matrix. Each column of this matrix is the word2vec semantic encoding of a sentence in our data set. We normalize each column to make sure that the magnitude of each column is the same, regardless of the number of words in each sentence. This allows us to perform queries simply by performing the dot product of each column with the query vector. The sentences with the largest inner product scores have the strongest semantic match to the query. We display a certain number of these sentences, which we call the responses to the interviewer's question.

Semantic querying uses word2vec embeddings trained on two corpuses or data: GoogleNews and LiveJournal. Each embedding is trained with the same master dictionary, to guarantee that each index corresponds to the same word. Let **w2vMat** $\in$ $\mathbf{R}^{\mathbf{300 \times \#words}}$ be the word2vec embedding, where **#words** is the number of words in the master dictionary, and 300 is the chosen vector dimension.

The featurized sentence data is the bag of words matrix described above. We call it **dataMat** $\in \mathbf{R}^{\mathbf{\#words \times \#sents}}$, where **#sents** is the number of sentences considered from the LiveJournal dataset. The sentence data is converted into vectors:

$$\mathbf{magic = w2vMat * dataMat,} \quad (1)$$

where matrix magic is further normalized along each column, and represents the numerical matrix wherewe perform the querying.

Querying if performed by converting the original query into a bag of words (**queryWords** $\in \mathbf{R}^{\#\mathbf{words} \times 1}$), and then calculating its corresponding vector

$$\mathbf{queryVec} = \mathbf{w2vMat} * \mathbf{queryWords}, \quad (2)$$

which is then normalized.

The semantic relationship is determined by cosine distance, which is the dot product of **queryVec** with the columns magic, since they are both normalized. The following array, **distMat** $\in \mathbf{R}^{1 \times \#\mathbf{sents}}$, represents the semantic score between the query vector and the sentence vectors from LiveJournal data:

$$\mathbf{distMat} = \mathbf{queryVec}^{\mathbf{T}} * \mathbf{magic}. \quad (3)$$

While the system above gives us valuable information, we will see in the following section that some additional features are helpful to tune the responses and obtain more qualitatively interesting responses. First, we use a minimum *threshold* on the number of words in the response. No response is allowed which contains fewer words than the threshold. All words, including stop words and out-of-dictionary words, are included in this total. We found that a minimum threshold of 7 is a good starting point for the LiveJournal dataset, but it can be selected by the interviewer at any time.

The interviewer is given the power to *filter* out responses which contain a specific word (or a word from a specific set). The tool does not return sentences containing one or more words from the set of filter words. One simple case in which this is helpful is to suppress responses with exact matches to words from the query question, thereby avoiding responses similar to those of a traditional keyword search.

We also give the interviewer the ability to add *importance weighting* to words in the search query. In this case, when calculating the embedding of the query sentence, we perform a weighted mean of the individual word embeddings based on their importance. Note that these weights can be positive or negative. In the latter case, responses with semantic relation to the negatively weighted word are suppressed.

In order to understand the context of answer sentences returned by the query, we also provide the interviewer a link to the original post on the LiveJournal website.

1.4 ITERATIVE CO-DESIGN PROCESS

Our population was qualitative researchers with diverse social science backgrounds. We interviewed

1.4.1 Initial Concept Formation

The design of this system starts with the realization that there is a potential to gather semantic data from large corpus of data, which could be used to help researchers identify what people are actually saying about a topic. The use of LiveJournal text helps access colloquial terms. The size of the corpus also ensures that even less popular topics are often represented in many journal posts. We also felt that existing keyword-based tools to explore such large corpuses are limited in their ability to recover unexpected insights.

We also want the researchers to be able to pose queries as full sentences in addition to collections of words. In this way queries can carry additional information such as emotions. We therefore formulated a very simple format of asking "questions" or queries to the crowd data to obtain relevant answers from the dataset. Figure 1.1 shows an example of a series of synthetic interview questions analogous to the semi-structured interview process. We expect that an interviewer hones their line of questioning along with their hypothesis by exploring the responses at each step.

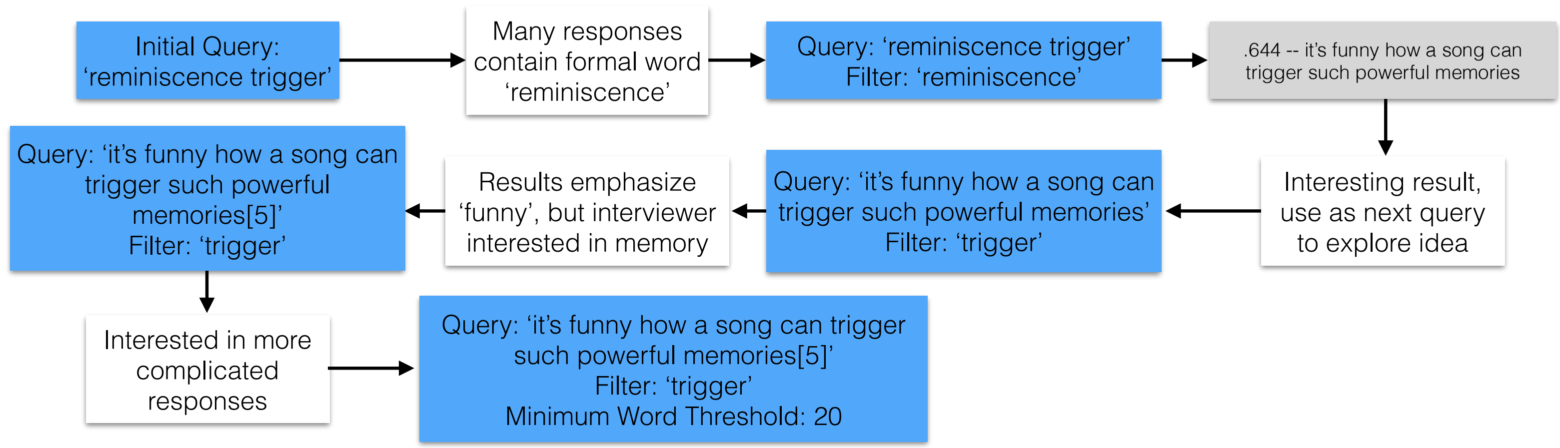


Figure 1.1: Example, based on a real query, o how the queries evolve in response to the returned answers. The flow is somewhat analogous to semi-structured interviews, where interviewers must modify their line of questioning on the fly to respond to answers received. The example is linear for simplicity of presentation, but the interviewer can easily follow a line of questioning and then return to an earlier query as many time as desired.

1.4.2 Early Idea Validation

Once we had established what we wanted to do we worked in two fronts, the method to obtain the data and the validation. We approached researchers very soon to gather perspective on the way semi-structured interviewing work and the overall perspective of the qualitative methods and tools. We started talking with a social worker and a couple of technology ethnographers, and simply asked them what they believed a large corpus of data could do for them. We used the analogy of giving them access to millions of diaries, and that they could "interview" their users through them. The social worker believed that such a tool would be very beneficial to have a "universal" perspective of people's thoughts. She believed that a fundamental need was to have "context", i.e. to be able to modify the unit of analysis from sentences to journals and back. The worker also mentioned that having a simple inter- face to pose queries is very important, especially if the tool can help search not only technical terms but also colloquial expressions.

> *If you can search in all those journals would be really good. . . and if there are many that are similar it helps to contrast.*

The ethnographers mentioned that their need to observe a population is fundamental, and a tool allowing them to describe a group of people with common interests would be valuable to help them when formulating initial hypotheses. One interesting conversation we had concerned an option to post-process the responses. Since some of our queries generated repetitive or redundant answers, one idea we had was to design a "diversity" filter to increase the entropy of our answers. Ethnographers believed that, in their field, "saturation" of responses (repeatedly receiving similar responses) could be informative.

> *For us seeing that something is repeated many times is valid. There is a concept in ethnography called "saturation", and we use it all the time*

They were interested in having such a feature if it could be toggled. They believed that a filter for diversity or even a simple option that allows them to see larger (more complex) expressions would help.

1.4.3 Low Fidelity (Lo-Fi) Prototype

Once we gathered this initial perspective we began to refine our search tool. Due to the complexities of the system, we had already developed an initial low-fidelity prototype that searched over just 0.1% of the data in one computer with 16GB of memory, where we could run queries on the fly. Later we progressed to a higher fidelity prototype querying approximately 1% of the data and we are currently working towards the ability to query 100% of the data on the fly using a cluster of computers. The reason we were so interested in using a prototype that will simulate the speed of the final system was that from our experience the speed at which the queries were generated was as important as the amount and quality of the content for creating a good search experience. This speed allowed interviewers to make quick connections between different queries and answers.

With this prototype working we presented the lo-fi prototype to two researchers. First we had a very short conversation with a public health researcher interested in the topic of diet, and how people maintain weight after dieting. He started by generating a couple of very simple queries about diet, and even though we were working with the lo-fi prototype he already found value in some of the responses. However, this researcher also immediately wanted access to contextual data such as the surrounding sentences or, even better, access to the sentence embedded in the actual post.

> *. . . if you can show me some context would be good. . . like, maybe the sentence before, or even better before and after. . . yeah, the link to the page would be good also, . . .*

During the process the interviewer asked if it was possible to weight the words in any way, as he noted that sometimes ancillary words dominated the answers. We took note of this as a potential improvement for the tool.

With this initial mildly positive reaction we carried on engaging other researchers and we already started to develop some way to access the contextual information surrounding the hit phrases.

A second researcher studies Airbnb and neighbor relationships with Airbnb landlords. The challenge in this case is the inherent difficulty in searching for a recent topic while our data set was obtained in May 2012 (we are in the process of obtaining a more recent version of the LJ corpus). Despite this limitation we explored queries concerning issues with neighbors. The frustration of the researcher was quite enlightening. At the same time, we observed a change in the query inputs, as the researcher moved away from the traditional keyword search approach and started searching more colloquial common expressions. Furthermore, she was interested in using the actual answers as queries themselves. Since we were having little success, this approach compensated our inability to frame a good query. This indeed brought us a bit closer to the type of answers the researcher valued. There was an improvement when searching on the mid-fi prototype (1% of the data) but, due to time constraints, we didn’t gather further insights.

Due to a time limitation we did a very quick search on the mid-fi prototype (1% of the data), and we did observe already an improvement. As closing remarks, the Airbnb researcher brought a very relevant point, which is that in order to generate good answers she believes that it was important for the interviewer to become a “smart user”. This was considered good and bad. On the one hand, the overall process of creating queries and seeing the answers did add to the knowledge and helped the researcher form ideas around the research topic. On the other hand, the perils and frustration of generating good queries should be addressed not only by improving the quality of the searches but also by adding a good UI to help creating the queries. The researcher also pointed out some tools like MaxQDA [86] as examples of some of the interfaces that qualitative researchers are used to, with good examples of knobs and other smaller helpful tools.

> *For the smart system vs smart user, I was just thinking along the lines of how well the system is equipped to predict what the user means (think of early Altavista vs Google today) and how that affects how easy it is for the user to get what they want – of course, the other side of this is how clear it is how the system works since if the rules of the game are obvious for the user, then it is easier to game the system to get what is sought for.*

Finally, we discussed the types of topics that LiveJournal can handle. The researcher believed that having some preliminary understanding of what topics are well represented in the corpus would help understand how adequate the corpus is.

> *. . . now, this is better, but you have to make some “crazy” assumptions to get to this level.*

One more interaction with a researcher focused on social relationships online. He was well versed in search models and he wanted to have an initial description of the system, which he understood well. As a matter of fact, the researcher had an inside story about a failed attempt by a commercial analysis software maker on trying to do semantic search a few years ago. He was very open to this idea and interested to see what he could find. He started with a simple query about “family holidays” (Figure 1.2 shows sample questions and responses related to this topic).

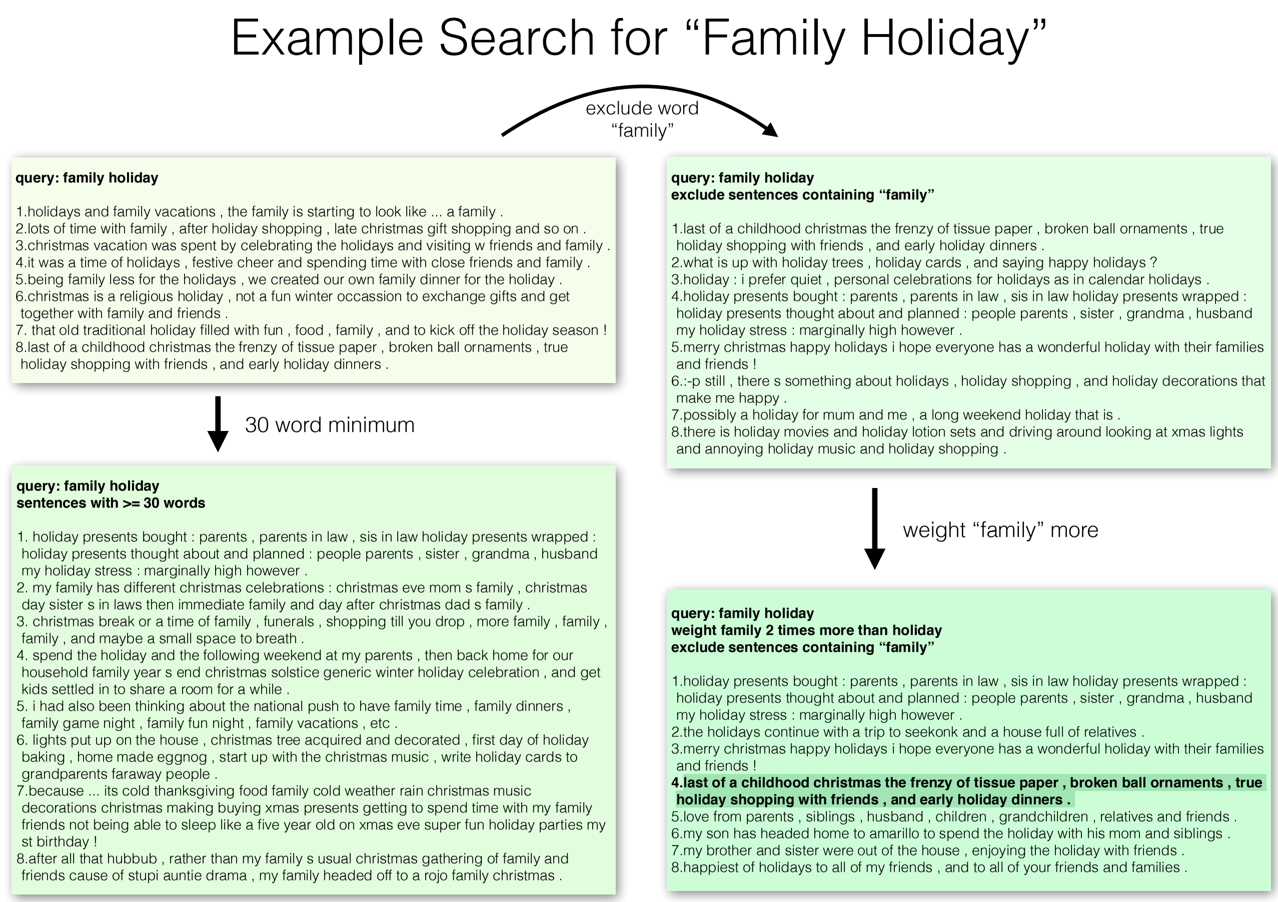


Figure 1.2: Example of the effect of filters in the query results on the full data. Top right displays the top 8 results for the query “family holiday”. To see more in depth responses, we can filter out the sentences with less than 30 words, see results in the bottom left. Since we perform a semantic query, we may choose to exclude the words “family” or “holiday” from the results. In the top right panel, we exclude the word “family”. Finally, we can define a weighting scheme on the query words. In the bottom right panel, we increase the weight on the word “family” to obtain results that relate more to “family” than “holiday”. Note that each result belongs to a full post, which we retrieve to the interviewer, e.g., the original post of the highlighted result is shown in Figure 1.3.

The social relationships researcher quickly found that the data was very promising. He was very happy with the large number of responses easily available, including expressions syntactically similar to the initial query as well as others semantically close, but quite different syntactically. We visited the actual journals as well (see figure 1.3 for an example), going back and forth between the response sentences and the full journal posts. He started finding meta relations across the sentences and we came across more than one answer from the same journal. The researcher felt that the synthetic interview tool was naturally helping him to find "genres" of users, and thought this was a great innovation.

MOOD: nostalgic

MUSIC: I'll Be Home for Christmas

Merry Christmas!

With Love,
From me to YOU!

Have a lovely Christmas holiday everyone. Eat food, drink, and be merry. And thus ends the Christmas season. This is my favorite time of the year. That fuzzy feeling inside warms the cockles of my heart, and yes, cockles is a word. Trees, mugs, smell of pine, songs, wrapping presents, lights, snow (or dreaming of a white Christmas), holiday movies, bells, bows, hot drinks, spirit, tradition, family, and love. What a cliched, corny holiday - and yes, I love it to death. This is the last you'll see of me prancing through the house in my underwearing singing Jingle Bell Rock at the top of my lungs, knitting to Miracle on 34th Street, and tipping everyone's cup with a splash of happy spirit. Come to think of it, this is the last *real* at home Christmas. It won't be the same when I come home from college, and the tree is up, the lights downtown are lit, and carols already sung and worn out. Last of a childhood Christmas - the frenzy of tissue paper, broken ball ornaments, true holiday shopping with friends, and early holiday dinners. Ah Christmas, I'll miss you.

On a lighter note, who wants to go after Christmas shopping? :0) I'm hitting department stores, shoe stores, malls...anyone can join in the fun chaos where everything is 50% and then 20% and then 10% off. Load up on those cards and wrapping papers, and leftover lights. Hit the old lady with a plastic reindeer as she trys to steal your clearance item. It's fun, trust me.

And for those who don't celebrate Christmas - have a wonderful day of rest and relaxation.

Figure 1.3: Public post on LiveJournal.com, corresponding to the highlighted result in Figure 1.2. By viewing the full text in its original context, the interviewer is able to explore a response in greater depth.

> *This is impressive...the actual responses for a query are very similar. . . Ah!, if you can search for all those "authors", people that write about the same topic you could get something like "genres". . . yeah, this would be very useful*

He believed that an ability to changing the unit of research from sentences to journals to genres of journals would be very valuable. The researcher also suggested that it might be interesting to store a history of queries, making a micro corpus from previous responses on which to perform additional semantic searches. He mentioned that, as the synthetic interview tool incorporates these additional complexities, some user interface (UI) tools would be necessary to make the process more tractable.

> *. . . exploring text is hard... yeah, some tools like remembering my searches or some faceted search would be good, but I do not want the system to make many decisions for me, it will lead us to a "search bubble"*

He was also very impressed that among the different results he was able to find phrases that were relevant to his query, but which did not include the actual query words at all. As you could see in Figure 2 the researcher searched for "family holidays" and found many results containing the word family, but also many that were about family without the actual word. He told us that he would like the ability to add a filter where the actual query words are excluded from the results.

> *. . . more than searching for the tail, it would be better to filter out "words" like this "family" word that comes many times and see only things like this one about the "brother and the sister"?*

Despite his remarks about UI tools, his biggest interest was indeed in this unique ability that our tool brought to potentially find "genres" of journals or groups of users, which for him was very impressive.

1.4.4 Testing on more data

By this time we had enabled a larger server with 64GB of memory to run the interactive queries on 1% of the data. We wanted to demonstrate this new iteration to the same public health researcher, who had a mild experience the first time we had interacted. The researcher began by asking demographic questions about the data set, such as number of users, posts per user, etc. He considered this important to frame his research. He also wanted to understand better the semantic searching process and asked whether it was better to search keywords or sentences. The researcher mentioned that he is very interested in finding users that are archetypes of those who used successful strategies to lose weight and keep it low. He had also prepared a group of synonyms about eating that he wanted to search. We started searching in the lo-fi prototype just as a way to see the difference. The public health researcher began by posing the question "i lost weight and kept it off" with no filter and the default threshold, but found little value in the results from the lo-fi dataset. We moved then to the mid-fi prototype, which return much more relevant results for the same query. He was very impressed and wanted to see various full journal entries. He felt that some journals were excellent case studies about his topic and he even thought of other potential research that people in public policy could make.

> *...and I think I will send this to my colleagues in the CDC, or even better get many other dissertations to other students in my department! :) . . .*

The public health researcher said that he would love to see some additional features like the ability to subtract words or to get the sentences before and after the matches. Finally, he said that he would like to be able to easily see the journals and select the ones he wants to work on and search on those journals alone. He also reiterated the importance of understanding the socio-demographic properties of the dataset. He was not concerned about knowing the topics embedded in the corpus, and not much interested in a predetermined diversity filter. He would not mind using it but only if he can control it.

1.4.5 Mid Fidelity (Mi-Fi) Prototype

In response to the feedback from the interviewers we implemented the additional search options (filtering, importance weighting, and minimum threshold) described Querying subsection above. We also expanded the interactive search to 1% of the dataset. We call this the mid-fidelity (mi-fi) prototype. We now wanted to contrast our former iteration with our filters and again contacted the social networks researcher and run some of his former queries but this time with filters.

He was delighted with the new results and made the remark that he believes that even at this level of fidelity he would be much more interested in using the tool.

> *...it is a really good tool to be able to find the chunks of data that I want to use in my studies and I would feel pretty comfortable in being able to use that [word2vec].*

Finally, we showed the tool to two other researchers. The first was a human computer interface researcher/designer studying online activism and multimedia communication tools for families. He found very little merit in the data for the activism topic, likely because his particular research topics were more modern than our dataset. Even with the mi-fi model, he did find some references that he found interesting once he adapted his search strategy from keywords to more colloquial expressions (such as “please sign my petition”). At this point he made a very important remark. He found that he had to completely change his search paradigm when using the synthetic interview tool, moving away from the example based queries that are successful with traditional keyword search engine queries and towards semantically meaningful questions.

> *I think the problem might be when you think about a search or a query tool we are way too much biased by our daily search experience, “all about keywords” but it seems to me that this, for this to work, you have to give an example.*

When he started searching for the information about a previous research topic, “reminiscence trigger,” the human computer interaction researcher was very happy with the results (see figure 1.4 for some examples). He said that the top results validated his past findings, from traditional research processes, that songs and smells are extremely triggers of reminiscence. He believed that synthetic interviewing has potential to support research at a formative phase as well as during refinement of the hypothesis. With more information about the population, the synthetic interview process could even become a useful formal tool.

Finally, we closed our interviews with a law sociologist re- searching access to justice. The initial search results were quite peculiar because of the formal phrasing of her queries. When searching for “presence in court” many answers were about tennis courts. However, when we added the filter to eliminate the word “ball” from the results the answers were much better and she found many interesting quotes about the judiciary system. She reiterated the point of previous researchers that access to detailed demographic information is a key element to make the tool viable. Without that it would be very hard to use it in publications.

> *I believe that I can use this tool in two moments. In the beginning of my research when I want to decide which issue I will talk about say I will try to figure out what is going on, what people are talking about, what are the main issues that are going on. I can bring one issue that is interesting to help me make the right question. It’s nice to see what people are talking [about]. I can then build the question to be tested. And then I see another time to use it when I want to make real research with people and I don’t have enough money or have enough time to go to the field, it will make it easier and more applicable to make research with people.*

Example Results

query: reminiscence trigger
exclude sentences containing “reminiscence”

1.but , just incase someone out there s supper trigger sensitive , trigger warning .
2.today ; s prompt , talk about a memory triggered by a particular song ?
3.things trigger off memories , memories that just sends me into pure rage , like this .
4.some smells not only trigger memories , they trigger the intense feelings that went along with them too .
5.this wireless trigger is a control discreteness for camera to trigger studio flashlight synchronously .
6.its funny how much a smell or a song can trigger such vivid memories .
7.my heart still melts at the constant reminders that trigger a series of memories .
8.is there a photograph or a song , for example , that triggers nostalgic memory of a certain period of your life ?

query: I lost weight by trying calorie counting
exclude sentences containing “calorie”

1.chris , the fat son , was trying to lose weight by dieting and exercising .
2.after dieting , exercising and trying to lose weight , i ve gained weight .
3.i have spent the past year , losing weight on a low carb diet .
4.i m back on the diet train , trying to lose some more weight .
5.speaking of weight , i ve already lost five pounds by drastically cutting carbs from my diet .
6.some of you may remember that i was part of the study comparing atkins weight loss to weight loss on a low fat , moderate carb diet .
7.and in my latest attempt to lose weight without exercising , i m drinking diet soda .
8.the last week i ve been on a diet , trying to lose weight .

query: democracy participation

1.political participation an educated , questioning , and engaged citizenry is essential for successful democracy .
2.a democracy needs the three fundamental building blocks popular sovereignty , political equality and political liberty to be a democracy at all .
3.bahamamama : democracy : nope : its a republic not a democracy ... who knew ?
4.trend of naisbitt s book , representative democracy to participatory democracy , continues the decentralization theme of the past two chapters .
5.it goes thusly : grassroots democracy is an organization composed of democratic activists and others determined to revitalize the democratic party .
6.harding the future of democracy another section from a citizen s guide to democracy inaction .
7.in his discussion of democracy , he cautioned that a pure democracy would oppress minorities .
8.social democracy your nation s freedoms : civil rights , economy , and political .

query: it's funny how a song can trigger such powerful memories
exclude sentences containing “trigger”
“memories” weighted 5 times more
sentences with >= 30 words

1.memories you wish you could forget , memories to remember , memories of the news , memories of the weather , memories of us , when we were together .
2.i keep having memories of her ... happy memories , sad memories , memories that fall somewhere in between .
3.that makes me sad ... choir banquet brings back memories , bad memories , good memories , and evil memories .
4.but next subject ... today was a day of memories , memories of jim hoff , memories of friends , memories of the summer .
5.thus , i have memories of remembering dreams , and sometimes memories of memories of dreams , but never memories of dreams .
6.she is the one who keeps the memories ; memories of tears running silent ; memories of threats issued harshly ; memories of pain too intense to explain ; memories of blood washed away .
7.has been a very long year , bad memories , sweet memories , frustating moments , stressed moments , happy memories , sadness , emo shit .
8.however , good memories make you think of more memories , and those memories make you think of more memories .

Figure 1.4: Example queries and results ran on 100% of the data, and with distinct filters.

However, the law sociologist believed that even without those key validation elements, the tool would be very helpful during the initial phase of the hypothesis formulation. She believed that the tool itself would be a great way to save money and time instead of going "door to door" at the early stages. She mentioned that for some of her projects, e.g. in the favelas of Rio de Janeiro, she was not even able to go to the field as it was too dangerous and she was forced to rely on secondary sources, like police or judiciary documents, to study the process of pacification within the favelas. She believed that if the synthetic interview tool could be applied to a corpus of data containing posts from such areas, this tool would allow her direct contact with the people's voice. Even in more accessible areas, she believes the simple ability to search quickly through what the people think about a topic is very compelling.

> *I had experience in Brazil because my field was impossible, so I decided to make research where I could have access available. So I decided to make a new social system. It was impossible to have people to classify the favelas, so I went to the judicial system database for the jurisprudence so I could make a research through the point of view of the institution.*

The law sociologist also expressed that she was eager to use the tool as soon as possible for her current research topic.

1.5 FUTURE WORK

This chapter was a first exploration of semantic matching methods as a alternative/complement to semi-structured interviewing in user studies. At high-level, we found that using semantic matching improved over traditional keyword-based methods, and that searching a larger dataset improves the quality and diversity of retrieved posts. These findings suggest a number of avenues for future work:

Improved Semantic Technologies. Word2vec is a fast and moderately effective semantic embedding scheme. But it ignores word order and phrase structure. State-of-the-art methods use recurrent neural nets (RNNs) such as LSTM (Long Short-Term Memory) [72]. These better model local and global sentence structure, and can be used to model paragraph and document-level structure. Using better embedding methods will likely lead to more useful results from semantic matching.

Larger-Scale Live Matching. Semantic matching is expensive at the full scale of Livejournal. With 500 million sentences and 300-dimensional embedding, one needs 600 GB of "hot" data in memory to do model matching. That requires either custom hardware or a cluster of machines. In the former case, one also needs a fast index to retrieve nearby sentences. Both solutions are practical but require integration of state-of-the-art techniques, and remain work in progress for us.

Synthetic Text Generation. One powerful affordance of the LSTM design is the generation of text as well as matching. That is, one can produce fully synthetic user output in response to a query. This synthetic text has high quality on related tasks such as language translation. While one loses the affordances related to exploring a particular user's post in context, one gains affordances related to the diversity and representativeness of the synthetic posts (via controls on the diversity of their synthesis). One also gains a level of privacy protection relative to retrieving true posts. The quality and utility of these posts is very much an open question.

1.6 CONCLUSION

In this chapter, we presented a new tool to perform synthetic interviews on a large, colloquial text corpus. We take advantage of advances in semantic word embeddings to provide a powerful semantic search tool based on queries phrased as sentences. Additional features for modifying and augmenting the search results were designed with the feedback of several qualitative researchers. The feedback from researchers indicated that the tool can serve some of the same goals as semi-structured interviews. The responses also highlight the tool's potential advantages in economy, interviewer reflection, representativeness, and interview control.

The synthetic interview tool bridges the divide between formal research questions and a large collection of informal anecdotes. As stated by the online activism researcher, this tool is unlike others in that it is "not all about keywords". Instead, the researcher states that he "can also profit from [our tool]" by simply "thinking of examples to search on". Compared to traditional keyword searching, the synthetic interview gives more varied results, including many whose connection to the original query is not obvious to the interviewer at the beginning of the process.

While a majority of our trials yielded positive results, we cannot ignore the remaining minority that did not echo similar success. One researcher voiced that because "the corpus is LiveJournal, [most authors] post about topics such as food, love, and marriage", however there "would not be [entries] on topics like physics". This shortcoming is in the query corpus, which we can target in two ways. In addition to our lo-fi and mi-fi prototypes described above, we were able to query 100% of the dataset offline (the process is currently too slow for interactive interviews). The results in Figures 1.2 and 1.4 are from such full corpus. This significantly improves the quality of many queries. We also expect that synthetic interviews on other topics which are poorly represented in the LiveJournal posts can be executed successfully by choosing an alternate corpus with greater relevance.

Another variable in the interviews is the choice of word embed- dings. We experimented with embeddings trained on Google News as well as LiveJournal posts, and concluded that the training corpus provides an additional option for tuning the synthetic interviews to specific research topics.

Chapter 2
Game & Narrative Design for Behavior Change

This chapter presents a list of principles that could be used to conceptualize games and narratives for behavior change. These principles are derived from lessons learned after teaching two design-centered courses around Gaming and Narrative Technologies for Health Behavior Change. Course sessions were designed to create many rapid prototypes based on specific topics from behavior change theory coupled with iterative human-centered and games design techniques. The design task was composed of two broad goals: 1) designing efficacious technologies, with an emphasis on short-term behavior change and 2) using metaphors, dramatic arcs and game dynamics as vehicles for increased engagement and long-term sustained change. Some example prototypes resulting from this design approach are presented.

2.1 INTRODUCTION

Persuasive technologies such phone apps or serious games share common goals of creating an engaging and efficacious experience towards behavior change. Either by modifying or adapting current interventions or by designing new applications based on behavior change theory, these techniques have the potential to reach millions of people who can benefit from a pervasive medium. Most designers and HCI professionals deal on regular basis with apps focused mostly on usability and engagement. In parallel, many health and biomed researchers focus mostly on high efficacy. Appropriate usability design may not be sufficient to guarantee long-term engagement - many times needed to gain adequate efficacy levels - however it is a needed condition for initial engagement/adoption. In this chapter we present a preliminary list of principles for conceptualization of games for behavior change derived from key lessons learned after teaching two semesters of design-oriented classes, focused on games to improve wellbeing and health.

2.2 PREVIOUS WORK

Several studies have shown that CCBT (Computerized Cognitive Behavioral Therapy) such as MoodGym [98], Beating the Blues [11], among others, compare very well with face-to-face therapy. However, engagement and attrition levels are not acceptable. Indeed, even though 2/3 of depressed patients say they would prefer therapy over drug treatment, only 20% of patients referred for in-person psychotherapy actually start it, and 1/3 of those will drop out [95]. Web-based therapy also has very poor engagement, although apparently for different reasons [84]. Dropouts may be due to (i) lack of commitment by patients (ii) lack of a regular schedule for system use (iii) difficulty or tediousness of using the tools.

Gaming has important characteristics that enhance some cognitive elements such selective attention [48], which can play an important role to behavior change, as they could help people pay more attention to the main message. Another very important characteristic gaming offers is that it also makes the learning process fun [83], which in turn generate better engagement. Complementary, games also help increase motivation [128] and emotional engagement. [59]

Previous work form Baranoski, et. al. [10] has already shown success using games for health behavior change. Some other gaming examples focused on health are the Personal Investigator [29] leveraging CBT for mental health, and Superbetter.us [136] which leverages real-life social support embedded into a superhero story.

Complementary to the gaming literature, the use of persuasive technology to improve usability and engagement for physical activity has been studied with the Ubifit system [27], which showed increased exercise levels by improving goal tracking, as well as using metaphors to improve people's engagement. Many other examples around exercise, sleep and stress reduction, such as Nike Plus [107], HearthMath [53] and FitBit [41] seem to indicate that systems associated with a lifestyle change have also higher levels of engagement among their niche adopters. In any case, it is yet to be seen if these technologies are set to be adopted widely.

the use of games to improve engagement we find the Lumosity [79] suite of games used for cognitive training. Cognitive techniques are wrapped around mini games, improving engagement, ensuring improved efficacy over time [129].

2.3 THE CHALLENGE: EFFICACY + EFFICIENCY

The challenge to merge efficacy and engagement can be dissected into the following design dualities (Figure 2.1):

Scientific vs. Iterative methods: A gap exists between current clinical intervention development methods based on the scientific method (hypotheses + statistical validation) and iterative gaming and app technology design. Usability design demands an approach that favors exploring ideas based on prompt user feedback through the construction of prototypes. However, it is necessary to keep in mind that the overall goal is to generate efficacious behavioral change that helps overcome health or wellbeing problems. Merging these two approaches is one of the constraints used to design our course sessions.

Short vs. Long-term focus: Short vs. long-term change is treated differently from a behavioral perspective. The former demands knowledge around decision-making, emotional elements and personal skills, while the latter demands a deeper understanding of identity and personality. A good way to mix behavior change goals with identity and personalization are narratives and games. These two elements incorporate concrete micro tasks associated with roles and missions that can be translated into smaller behavior change skills, while the metaphors, scenarios and stories support a deeper immersion into new identities.

Content vs. Dynamics: When designing interventional technology, efficacy is usually regarded as the main goal. Engagement usually plays a secondary role, which could have a major impact in the adoption of the technology. Commercial apps do look for a more complete user experience, which pays attention to execution as well as engagement and identity details. However, success is usually measured in terms of revenue generation, rather than behavioral metrics. Merging both the content (i.e. narrative) as well as the dynamics of the game into a coherent design that help develop real life skills is yet another design challenge to be considered.

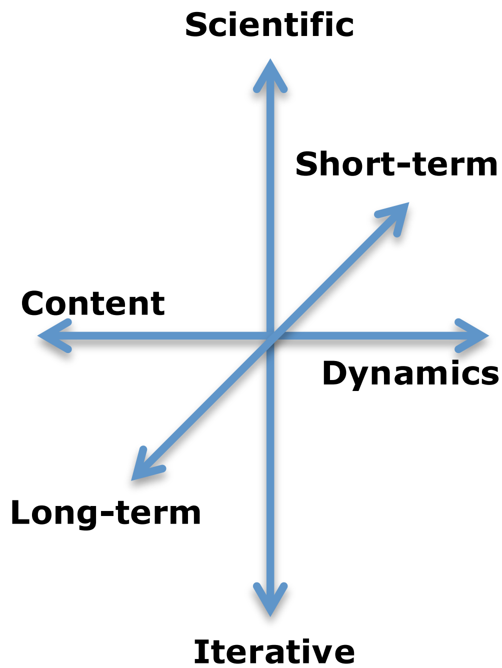


Figure 2.1: Efficacy + Engagement game design

2.4 DESIGN METHODOLOGY

It is important to note that the methodology followed during the conceptualization process in the course is focused on maximizing creativity provided the aforementioned constraints. To help students acquire sensibility around behavioral efficacy, specific behavioral theories are used as the basis of design challenges to promote rapid prototyping in very short sessions. A specialist, who many times has little design experience, presents the theoretical component in a one-hour talk. Students can ask questions associated with the topic being presented, and the specialist intervention closes with a brief discussion around the way such behavior change theories can be used to design new technology. Table 2.1 shows a list of the theoretical topics presented. During the second hour a design challenge is presented to the students. They need to go from problem assessment to a complete game concept with rules, usability scenarios, title and introduction. In many cases we even ask them to create a suggestive

billboard to position the idea. Students need to begin by expressing a behavioral problem through its disempowering narrative and find the counteracting empowering narrative, which leverages the behavior change concept taught by the specialist. The students must storyboard both narratives (disempowering and empowering). They must also externalize relevant intangible elements such as the problems themselves and/or the feelings associated with it by converting them into enemies, scenes, obstacles or other gaming elements. Additionally, they need to externalize the skills needed to overcome such problems by portraying them as weapons or as specific game dynamics. At all times, students are encouraged to make sure game progression is elicited and not only end goals, to make sure change is embraced by the users. Finally, students are asked to test each other's games, present their game as if it was being launched on TV or act their games out.

Behavior Change Topics	**Narrative Topics**
Intro to Behavior Change	Intro to Life Stories
Body-Mind Connection	Narrative Psychology
Positive Psychology	Drama Therapy
Sports Psychology	Neuroscience Games
Anxiety, Depression and Cognitive Behavioral Therapy	Trauma Narratives
Behavior Change in Society	Improv-based Games
Communitarian Mental Health Interventions	Digital Storytelling
Social Networking for Behavior Change	

Table 2.1: Behavior Change and Narrative topics taught in addition to Game Development and Human Centered Design topics.

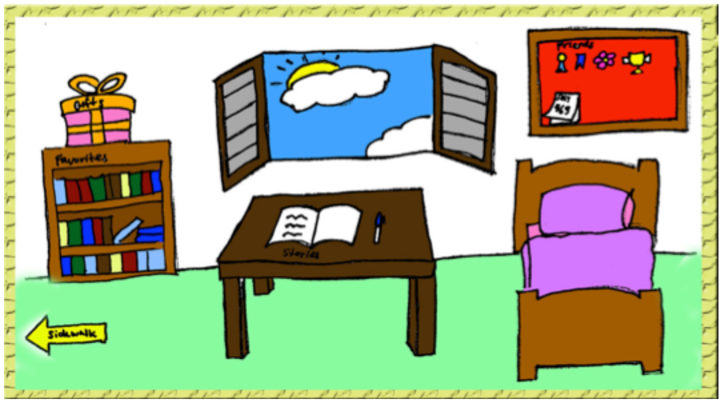
2.5 GAME PROTOTYPE EXAMPLES

Among many others we chose a few examples of the work done in class:

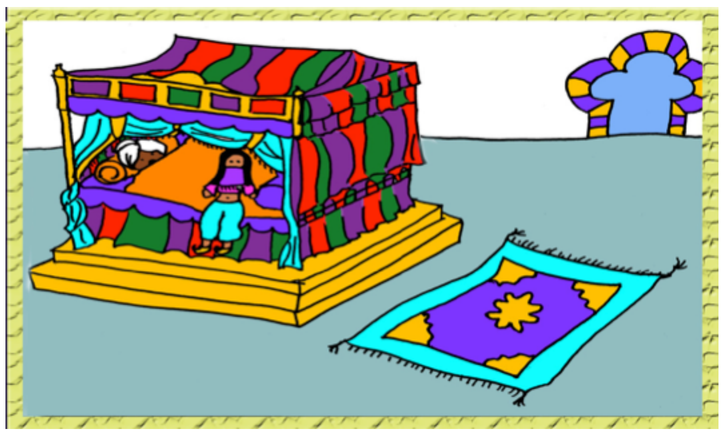
Monsters (Figure 2.2) – a simple two-player game based on monsters and weapons**.** Concepts around externalization and empowering metaphors drawn from drama therapy and narrative therapy are used to make "visible" enemies and the weapons to destroy them. These elements have clear links to the problems and skills needed to solve them. For example, a monster representing stress can be seen as a flaming monster, and player 2 can be used to help you blow the torch by teaching you how to breath correctly to calm you down.

Figure 2.2: Monsters Game Screen

Scheherazade's World (Figure 2.3) – a game that aids in the prevention of suicide by creating a community for at-risk young women to share their stories. The One Thousand and One Nights tells of a king named Shahryar, who would marry a new wife each day and sentence yesterday's wife to death. Unlike previous wives, Scheherazade had a secret weapon to keep her alive. Every night, she would tell the king a story, only to end with a cliffhanger each night. Because the king wanted to know the rest of the story, he would spare her life for another day. Through stories, she was able to survive. This game is based in part on Narrative Therapy and Drama Therapy aspects, as well as Digital Storytelling.



Bedroom – Write stories here



Sultan's bedroom – Listen to Scheherazade

Figure 2.3: Scheherazade's Word Game

Semester Adventure (Figure 2.4) – a simple game to reduce stress and improve time management around test exams, where the player follows an adventure as a warrior that needs to reduce stress by gaining powerful tokens by improving his/her time management skills, i.e. fulfilling tasks on time, which are portrayed as enemies to be beaten. This game leverages personality theories based on life stories, which indicates that people assume new roles based on the way they define themselves. A "warrior" narrative helps people to be active and assertive, while a "victim" makes the person passive and receptive of disgrace.

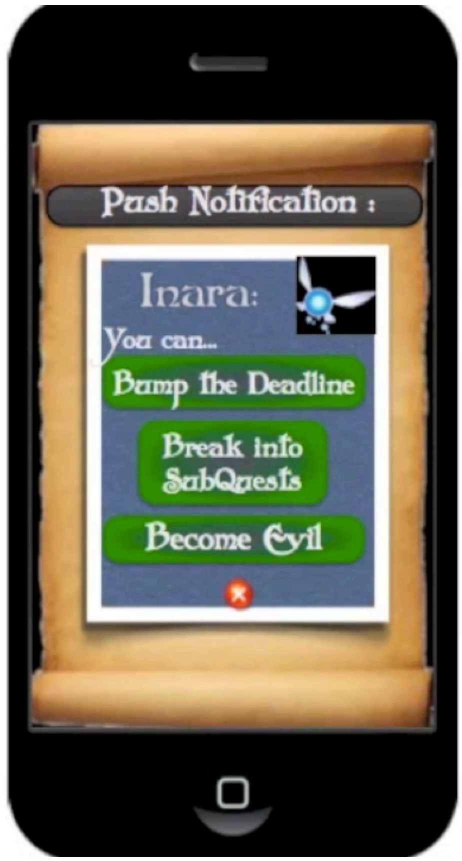


Figure 2.4: Semester adventure game

2.6 NARRATIVE PROTOTYPE EXAMPLE

2.6.1 Mental Health Background

The World Health Organization reported that mental disorders are a major financial burden for the world's economy. They are currently the third most costly health problem in terms of disability-adjusted life-years [105] around the world, and the largest in the US and Canada. They represent 10% of the global disease burden. Fewer than 25% of depressed patients receive the necessary treatment and an even smaller percentage of at-risk groups have access to adequate preventive care [105]. In many developing countries, mental illness is not even recognized as a medical problem [94]. Furthermore, (untreated) depression rates are very high for chronic disease patients in those countries (e.g. AIDS patients) and impede patient's treatment, compounding those health problems [94]. A multitude of barriers exist, including a lack of trained professionals and understanding of mental illness by patients themselves. Even in developed countries, mental illness is stigmatized and many people will not seek treatment.

In this section we introduce a machinima (machine + cinema) system that will empower therapists, social workers, public health professionals, to further increase the efficacy of therapy. Machinima is an animation recorded from a game engine where a human "directs" the actions of the characters. For our application we have focused on the SIMS engine because of its rich emotive content already available, in fact, most of the scenarios and characters required were readily available. Machinima allows the creation of engaging entertaining videos as an adjunct to, and as a potential alternative to, existing approaches. CBT is an effective approach for a variety of mental health concerns including depression, anxiety, insomnia, eating pathology, marital stress, and chronic pain [21]. CBT principles have been integrated into several computerized formats as both an adjunct to face-to-face therapy and as a stand alone treatment. Computerized CBT (CCBT) is a fully automated implementation of CBT where patients follow a complete treatment through interactive texts and figures. Several studies have shown that CCBT compares very well with face-to-face therapy [71, 95], and CCBT is approved for reimbursement under Britain's National Health Service (NHS).

Indeed, although therapy is effective and patients prefer it to psychopharmacological treatment, only20% of patients referred for in-person psychotherapy actually start it, and 1/3 of those will drop out [84]. Thus even in developed countries, there are great barriers (i.e., stigma, lack of access, anxiety about live discussion, etc.) to conventional treatments for mental health problems. Web-based therapy also has poor uptake, although apparently for different reasons [100]. Dropouts may be due to *(i) lack of commitment by patients (ii) lack of a regular schedule for system use (iii) difficulty or tediousness of using the tools.*

Our approach uses scripted health interventions following the design principles of a number of successful systems, but adding two incentives: (i) use of an *interactive entertainment scenario format, and (ii) engagement of family members in the therapy process.* Although we focus our initial efforts in the treatment of depression, we believe the values of a cinematographic approach to engage patients who could benefit from CBT in general. To test our idea, we are currently producing the material that translates the CBT-based manuals for the treatment and prevention depression developed at San Francisco General Hospital [100]. Figure 2.5 shows both; the printed material that is distributed to patients, and some of the machinima screenshots of the micro video created to enhance/replace this material.

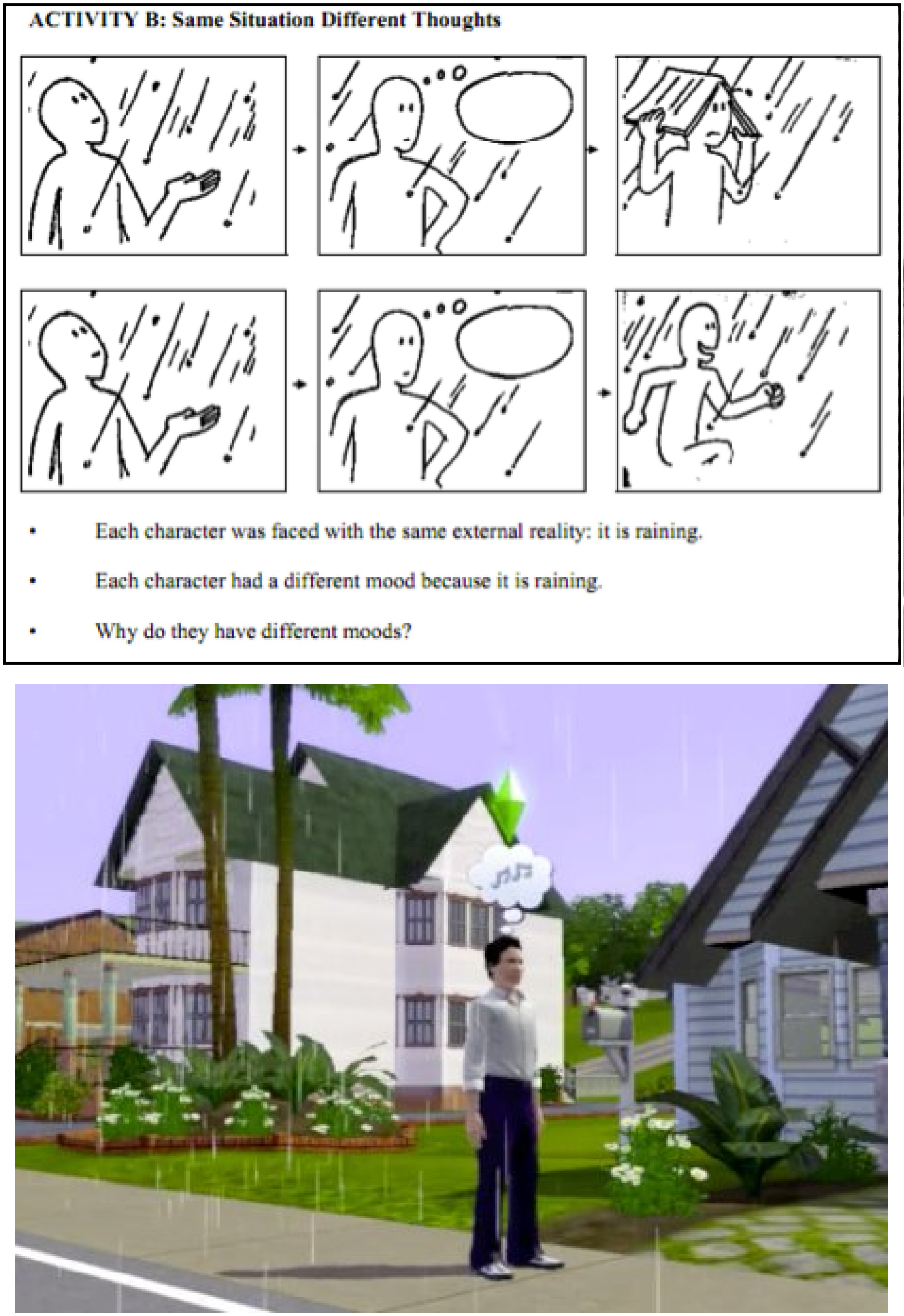


Figure 2.5: Original CBT cartoon versus machinima video screenshot.

2.6.2 Related Work

Among the CCBT commercial systems available that contain interactive cartoon elements the most relevant are: MoodGYM [98], FearFighter [39], Beating the Blues [11], Living Life to the Full [76]. These systems are endorsed by several governments and provide effective treatment even through insurance coverage in some cases.

In contrast with current efforts focused on computer games for CBT enhancement (Coyle, et. al.) [30], machinima could drive adoption of self-help interventions because a game engine not required. Although machinima is derived from games, it does not require a game engine. Rather it is a series of modest-resolution videos with decision branching. It is modest in both memory and CPU usage. These modest requirements could support devices with limited interactivity such as DVDs, which in turn can drive higher adoption in low-income populations

2.6.3 Prototype Movie: "A Rainy Day"

We created a prototype based on the San Francisco General Hospital Depression Clinic manual. This manual is currently used for group CBT of major depression and maintained by the Latino Mental Health Research Program team from the University of California at San Francisco [100]. The manual consists of four modules (thoughts, activities, people and health) covering key CBT concepts and content relevant to primary care patients. For our prototype we drew from an activity included in the first module (Thoughts and Your Mood). Specifically, we adapted a cartoon that demonstrates the link between thoughts and mood. Figure 2.4 showed both the cartoon and some movie shots. In this simple sketch, a character walking reacts to the beginning of rain. In the first part the character chooses an unhealthy thought and reacts with sadness. In the second part, the same character has a different thought that triggers smiling and leads the character to run happily through the rain.

Figure 2.6 shows the way the information in the original manual was converted into a movie script. CBT training material was translated into a movie script that followed basic rules of cinematography, which described the scenarios, camera shots, actors, dialogue and transitions.

COGNITIVE BEHAVIORAL TREATMENT FOR DEPRESSION
PARTICIPANT NOTES: Thoughts Module: Session 1
Version:May, 2000

EXTERNAL AND INTERNAL REALITY ARE BOTH IMPORTANT

External/Objective Reality - The facts: parts of your reality that are observable and measurable.

- the things you do
- illnesses you have experienced
- how much money you have
- how many people live with you
- your physical surroundings

Key point: although your external reality may seem fixed, parts of it are changeable. For example, you can decide where you spend your time. You decide whether you stay inside or go for a walk. There are parts of your external reality that you can manage. You can choose to be in those parts of your external reality that are helpful and healthy for you.

Internal/Subjective Reality - The world of your mind, which is yours: not observable by others.

- thoughts
- memories
- beliefs
- expectations
- the way we understand what has happened to us.

Key point: You can change and manage your internal reality. You decide which aspects of your reality you focus on. Changes in your external reality will affect your internal reality. By changing a part of your external reality, you can change future memories, beliefs, and expectations.

You cannot completely control either your external or internal reality. Both your external and internal reality are real. Both are important and both affect each other constantly. But it is important to remember that some part of each one can be changed.

Depression is not all in your head.
Depression is affected by what you do, how you think,
and what happens to you
And how you react to what happens to you.

Healthy Management of Reality – Machinima Script
7.18.2011 – Paredes/Schueller

Plot/Action: Thought + Behavior responses to real events.
Subplot/Theme: Rain - As a trigger of thoughts/behaviors and as a reminder that you can control them.

A Rainy Day

Fade In:

EXT – Suburb park – Day

It is a sunny day in a suburban park location

Camera approaches park slowly to get character in focus.

NARRATOR:
THOUGHTS ARE PART OF OUR INTERNAL REALITY. WE PROCESS WHAT HAPPENS TO US WITH OUR THOUGHTS. THESE THOUGHTS EFFECT HOW WE FEEL, EVEN GIVEN THE SAME SITUATION. THE SAME EXTERNAL REALITY.

Character with a backpack/case is walking down a street placidly. Suddenly, it starts to rain.

Camera focuses on rain falling on the ground or tree.

Character pauses, holds his hand out, noticing the rain drops. Character pauses, thought bubble appears. Character begins walking with book covering his head, looks dejected.

Camera closes up on subject's face.

NARRATOR:
WHAT DO YOU THINK HE WAS THINKING? HOW DID HE FEEL AFTER THAT?

Camera focuses on rain falling on the floor or tree.

Fade Out

Fade In:

NARRATOR:
NOW CONSIDER THE SAME RAINY DAY!

Character with backpack/case is walking down a street placidly. Suddenly, it starts to rain.

Camera focuses on rain falling on the ground or tree.

Figure 2.6: Excerpts from original CBT manual and machinima movie script.

We created the video using several cinematographic elements available in machinima: background music, foreground texts, a narrator, camera panning and scenery. We were able to further introduce subtle yet powerful elements that transmit strong emotional feelings, which should help the user remember this lesson by enhancing his/her memory associated with this emotional arousal [22]. We used camera panning to have close-up views of the character, which allows emphasizing the feeling of despair felt by the character on the first part of the video. On the contrary, an open camera shot provides a feeling of realization, freedom, and completeness. We used the rain as our anchor element to elicit a theme centered on thoughts. The rain was the trigger that remained invariant in both parts of the movie. By focusing on the rain drops when it starts to rain, and also after the character has expressed his thought and mood, we elicit the theme of rain as a neutral event that triggers a thought that is entirely dependent on our interpretation of the situation.

Machinima's cinematographic and narrative elements (camera angles, sounds, music, plot and subplot, characters), the person's sense of immersion in the scene and its verisimilitude to real social experience, should further increase not only the interest of the patient for the video (Table 2.2). For example, the rain theme can be understood as the emotional trigger and as the element that defines our power to control our own thoughts, making it an element that could be easily remembered.

Elements	Detail	CBT value	Entertainment Value
Title	A Rainy Day	Focus on reality	Catchy phrase to remember
Plot (Action)	Thought Selection	Building block for CBT therapy	Visual thought blobs that explain and elicit thought process
Subplot (Theme)	Rain	Thought trigger	Emotive engagement through a conflict to be resolved
Protagonist	Simple person	Resemblance of reality	Realistic avatar catches our curious attention and immerses us in the plot/subplot
Scenery	Simple park	We can relate to a place like this	Beautiful surrounding to be challenged by rain
Narrative/ Dialogue	Narrator	Voice of the instructor	Calming voice that provides hope and explanations

Table 2.2: Movie script elements of "A Rainy Day"

In summary, CBT's engagement and efficacy issues can be improved by easily adding a mix of entertainment, narrative and dialogical interactivity.

2.6.4 Uses of Therapeutic Machinima

Machinima’s flexibility should allow for the creation of databases of videos with minor modifications that help tailor the best content to the user. Generic storylines with small changes in locations (park, street, city, etc.), characters (sex, age, race, etc.) will allow for health practitioners, coaches, educators and other members to benefit from a rich resource tailored to important places or situations in a person’s life.

It is clear to observe that most of the advantages that machinima adds to CBT can also be applied to other types of training material. Therefore, we believe that machinima has the potential to push for a dramatic increase in the creation of educational and training systems leveraging highly customizable video.

Complementary to the research and technical challenges related to machinima, the use of video should be proven effective to improve engagement of illiterate users, and therefore use this technology to help breach deficiencies in mental health treatment in marginalized regions.

2.6.5 Future Research

Machinima’s impact in the creation of tailored content for mental health must be evaluated in three aspects:

1) Is machinima more engaging (better persistence, better uptake) than non-game methods?,

2) Are there any improvements in efficacy from its higher level of interactivity and engagement?,

3) Can a reasonable degree of branching be added without creating an overwhelming number of videos to make?

Another research challenge would be to produce a series of videos where characters and scenarios can be designed for different family members to transmit different components of a family psycho-education program.

Research on the effects of perspective (first and third person views) should be performed to determine which perspectives are more appropriate to teach specific CBT concepts as well as the effect on the engagement of patients and family members.

2.7 PRINCIPLES FOR CONCEPTUALIZATION

2.7.1 Understanding Disempowering Narratives

a. *Narratives are lived* - not only used to tell stories about one self. People confront ideas and situations based on the way they portray themselves. This is observed in trauma patients who cannot overcome the generalization of their disempowering narratives. Understanding the narratives new empowering narratives underneath unhealthy behaviors will help design new empowering narratives that change unhealthy habits, eliminate over-generalizations and organize thoughts around the appropriate context.

b. *Focus on strengths* – Design around behavior change can benefit from understanding people's current strengths, rather than imposing an ideal model for functioning under a specific situation. A key concept that describes the basis for behavior change is what Bandura defines as self-efficacy [8]. In a nutshell, self-efficacy explains using current strengths. However, discovering strengths may demand an exploration not only of thriving experiences, but also difficult experiences, where strengths are used to be resilient and survive emotional or physical pain or disgrace.

2.7.2 Externalizing Problems

c. *Interpretation and introspection* – Problems are rarely completely understood by users. Designers should strive to provide tools, time spaces and cues to help people interpret problems and introspect. Games with forced pauses and prompts for reflection could help increase people's awareness of their own thoughts and further understand their problems. Furthermore, health behavior change games must be designed to be adaptive to changes in problem definition, as the game helps the user discover the root cause of a superficial problem.

d. *Problems as fictional enemies* – Externalizing problems into concrete game elements (i.e. objects, monsters, obstacles, etc.) help people understand that a problem does not occupy every aspect of their lives. It also helps the user understand the characteristics of the problem, which in turn will help understand the possible solutions around it. Designers should provide users the possibility to externalize their problematic feelings into a concrete game or narrative element that can later be destroyed or controlled. The element representing the problem should have a clean metaphor, for example, stress into an oppressive rock, or depression as glue that impedes you to move, in order to help the user understand the characteristics and affordances of the problem at hand.

e. *Materializing skills into weapons* – As well as tangible problems, weapons that represent the skills required to overcome the problem should be materialized. Such tools must carry a clean meaning that is memorable and supports the notion that change is possible via the use of the metaphors associated with such weapons.

2.7.3 Game Dynamics as Interventions

f. *Progress as a proxy for self-efficacy* – Eliciting progress should be a key element of game design for behavior change. Many times users need to realize first that "change" is actually possible. If no progression is clearly observed, the sensation of inefficacy is perpetuated and therefore, any additional effort to develop skills or change motivations could be futile. The initial game levels must demonstrate to the user that change is possible.

g. *Social validation* – Sharing and celebrating with others helps assimilate the new changing reality. Without social affirmation around change, progress may seem part of our imagination. Designers should use social affirmation to promote self- efficacy. Using social influence could be used as a vehicle to get some concrete change, but it runs the risk to leave the user believing that they were imposed a new reality by others and therefore reducing gains in self-efficacy, which ultimately drives change.

2.7.4 Narratives as Interventions

h. Translate abstract concepts to visual stories: CBT's lessons learned by patients are mostly abstract. Core concepts of CBT include the investigation of the link between thoughts, mood, and behavior and the challenging and reframing of thoughts to improve one's mood and promote healthy behaviors. These concepts are complex and are made more concrete via examples and metaphors. These examples can be better presented with graphics or even better, with storyboards or cartoon elements. In the case current commercial CCBT systems [39, 10, 76, 98], these examples are presented as interactive cartoons that allow the user to make choices along with the storyline.

i. Verisimilitude to real social experience: With machinima we add a much greater level of realism: facial animation, body gesture, movement and speech prosody, with a tinge of playfulness and irony inherited from the source game (In this case the SIMS). The game mechanics available in the SIMS, allow easy setup cameras, control and program goals for avatars that already contain high resemblance with human actors, including emotional content. These mechanics provide flexibility for authoring from the perspective of a producer or director, rather than having to develop complete scenarios and simulate human characters.

j. Simple game mechanics: CBT's main driver of progress is through homework that reinforces skills taught in sessions and allows people to apply these skills in their daily lives to better promote healthy thought management and behavior change. Many of these tasks are guided through printed materials, or through notes taken by the patient during its therapy sessions. CCBT adds pictures and online questionnaires that adapt to the patient progress to further enhance engagement and retention. Narrative games (machinima) should provide an element of interactive fun that will engage the patient in a dramatic plot while learning lessons. Additionally, this element of cinematographic fun could be shared with family members to increase awareness or even support the learning of some interpersonal skills.

k. Continuous and adaptive learning: Principles either learned at therapy session or self-taught can be reinforced through constant reviewing of the material (through DVD, mobile phones), and are made more salient when accessible to patients after an event evokes the relevant lesson. Although these concepts could also be replicated in paper-based materials, the interactive nature of machinima elements allow the user to follow a decision tree that can adapt to the learning and understanding progress of the user.

l. Family engagement: Videos can be designed into a mini series that follows a family psycho-education[83] program aimed to: 1) generate understanding of the patient's problem, 2) incorporate different perspectives to similar situations to generate empathy, 3) complement the patient's education by explaining hard to grasp concepts that must be discussed in family, 4) social capital is restored by sharing entertainment and educational moments, 5) allowing the patient to re-experience techniques learned in session with the therapist.

2.8 CONCLUSION

In this chapter we present a framework to guide the way we conceptualize behavior change interventions. We propose a design space that takes advantage of game and narrative mechanics and content, that describes a short or long term strategy, and that leverages a scientific model for efficacy versus an iterative model for engagement. We present some prototype examples of game apps constructed based on theoretical basis. We present a narrative prototype that leverages cinematographic elements and machinima and we conclude with a series of principles for design that discuss issues of disempowering narratives, externalization, game dynamics and narrative dynamics.

Chapter 3
Sensor-less Sensing

This chapter describes our vision on what should be the research around sensing that enables adaptive interventions to make affective computing and stress management technology pervasive and unobtrusive. With the use of common computer peripherals and mobile computing devices as affect sensors, personalized and adaptive intervention technologies can be developed. Furthermore, physiological sensing can be performed without the introduction of extraneous factors such as wearable devices or focused software. Different methods for sensing and complementary adaptable interventions and interactions are described and proposed. We show some empirical evidence of the use of a computer mouse in the detection of stress.

3.1 INTRODUCTION

This chapter presents the body of knowledge around the adoption, usability and design of non-invasive sensing techniques for affective computing and stress management, due to its big impact in mental health and productivity. We define sensor-less sensing as the umbrella term covering the opportunistic use of existing sensors embedded in daily use computing devices and peripherals such as mice [55, 82, 135], keyboards [31, 37, 55], cameras [12, 118] and mobile devices to repurpose their signals to track different biometric states representative of mental or physiological states directly or indirectly. Sensor-less sensing offers a great opportunity to detect physiological metrics such as heart rate, heart rate variability (HRV), breathing rate, etc. Another exciting opportunity is to indirectly detect mental state changes. Preliminary research performed in our lab on this topic has been able to detect robust signals in voice [24], and by measuring the natural oscillation of the arm through a computer mouse [135]. In a first approach sensor-less sensing can be used to monitor affective states associated with the use of computer interfaces (such as frustration, anger, happiness, stress, etc.) There are a number of advantages in leveraging daily use computing devices as “sensors”:

Accessibility - peripherals such as mice and keyboard are ubiquitous and are indispensable to interaction with desktops graphic user interfaces.

Unobtrusiveness - There is no need for wiring sensors to the body or speak to a microphone (or cellphone) especially in semi-structured office spaces.

Long-term, in-situ monitoring - Many people spend a substantial amount of time using computers, which affords the opportunity to monitor and provide feedback while people are engaged in stressful tasks.

Application and content neutral monitoring - Mice motion and smartphone gestures is neutral to the application and the content with which the user is interacting. This implies better generalizability and alleviates privacy concerns compared with other more intrusive techniques, like monitoring keystrokes or camera usage.

3.2 BACKGROUND

3.2.1 Arousal and Stress

Arousal and stress affect almost all the body functions, including cognitive function and memory. Stress/arousal is not a single system but several related systems working in harmony (allostatic body function) [17]. Stress has a concrete function to trigger a reaction to a specific threat (stressor) [20]. Arousal generally improves memory, speed of work, association and pattern recognition; it also increases errors on unfamiliar tasks and causes cognitive and perceptual narrowing. The effects of these changes on performance depend on the task.

This general form has been observed in studies on the effects of emotional stress on recall [25]. However, a large body of research has given conflicting results under different stress/performance combinations, often with linear relationships between arousal and performance, or with performance that saturates at high levels of arousal. It is best to assume that this relationship needs to be learned for each class of tasks. Similarly, it has been shown that the optimal level of arousal varies from one individual to the next. This all suggests that in order to make use of stress/arousal feedback, a system needs to gather a large amount of data about each individual, and furthermore this data needs to be identified with the task the subject is doing.

Traditional psychophysiology focuses on the Autonomic Nervous System (ANS) due to its strong correlation with affect. Its quasi-independence from the Central Nervous System (CNS) helps to use it as a stronger signal associated with emotions. We wanted to take advantage of the less explored Somatic Nervous System (SoNS) . The use of movement from limbs as a proxy for arousal is intuitive. Changes in muscle tension and grip are associated with emotional change. However the challenge is to obtain a pure signal, rather than reading other mental processes. Attention, intention, workload, etc. are signals that affect SoNS. Methodology is important to isolate the proper signal.

3.2.2 Physilogical Affect Sensing

Recent advancement in sensors and wireless technologies and related computational techniques have accelerated the push towards wearable health-care devices capable of providing ambulatory monitoring of a variety of vital signs, such as electrocardiogram (ECG), electromyogram (EMG), pulse oximetry, bio-impedance, electro-dermal activity (EDA) (formerly known as Galvanic Skin Response - GSR) [106].

Many of these vital signs are strongly linked with physiological changes induced by emotional arousal [116]. Healey and Picard used a combination of these physiological sensors to determine stress-levels for drivers in real-life driving tasks [52]. However, many have also reported that a variety of other conditions such as physical activity, attention or fatigue can introduce physiological responses very similar to stress/arousal [58].

Affect sensing using voice and facial features has also been explored. Specifically, voice analysis has demonstrated modest accuracy at estimating emotion and high accuracy of stress/arousal [24]. Computer vision has demonstrated good accuracy at emotion detection from facial images [74]. However, collecting sufficient data to train robust models for voice analysis is challenging, and in both cases deployment is further complicated because of privacy concerns [77].

In spite of their growing availability, friction continues to exist in public adoption due to the intrusive nature of body periphery sensing. By enabling daily use computing devices capable of converting some body signals into mental health metrics, new affective technology adoption could be improved.

3.3 AROUSAL FROM COMPUTER PERIPHERALS

3.3.1 Mouse movement[1]

A promising exploration of the use of mouse movements to detect affect was Wolfgang Maehr's diploma thesis [82]. Maehr used several metrics on mouse movement, and emotions induced in subjects by watching short videos. Specific "motion breaks" that were discontinuities in mouse movement were significantly related to both arousal and disgust and close to significant for anger. One common correlate of arousal/stress is muscle tension. Tension in arm and wrist muscles would change the dynamics of the movement, e.g. its resonant frequency and damping ratio.

In our work (Sun, et. al. [135]) we have seen that it is possible to train simple controller models, such as Linear Predictive Coding (LPC), to capture the basic dynamic parameters of the movement. We approximated the parameters of a Mass-Spring-Damper (MSD) model. The mass are the bones and muscle mass, while the spring is the muscle tension. The exploration used "game grade" mice, which have update rates of 500 Hz and resolutions of 5700 dots per inch (DPI).

We used three different tasks: Clicking, Drag-and-Drop and Steering. We randomized four different distances and four different object sizes. Users performed all with or without stress. We calmed down users at the beginning and after each stress task. We measured movement only after a stressor (math) or a calming intervention (positive visualization).

Table 3.1 shows the results for the different tasks during the use of the mouse posterior to the calm (post_calm) and the stress (post_stress) tasks. We show only results for the x-axis. Under a t-test, we observe statistical significance - after Bonferroni correction - ($p<0.0167$) for all the tasks along the x-axis. For the Clicking task, the damped frequency was on average higher during the Stress phase than the Calm phase, along both x-axis (ωx) ($t(48) = 4.54$, $p<0.001$) and y-axis (ωy) ($t(48) = 3.94$, $p<0.001$); while

[1] Some of the data presented in this section is taken from: Sun, D., Paredes, P., & Canny, J. (2014, April). MouStress: detecting stress from mouse motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 61-70). ACM.

damping ratio was lower along the x-axis (ζx) ($t(48) = 4.54$, $p<0.001$). No significant effect was observed for the y axis. For the Steering task, the analysis showed that the damped frequency was higher during Stress compared to Calm for ωx ($t(48) = 2.40$, $p=.01$). In the y-axis only ωy was significantly different ($t(48) = 2.55$, $p=0.007$). No significant effect was observed for damping ratio along x-axis or y-axis. None of the parameters were significant for the Dragging task. Time was significantly lower under stress only for the clicking task ($t(48) = -2.65$, $p=0.005$).

Task: Clicking	**post_calm mean(stdev)**	**post_stress mean(stdev)**	**t(48)**	**p-value**
Damped frequency (ωx)	0.13(0.001)	0.14(0.001	4.54	<0.001*
Damped ratio (ζx)	0.53(0.0003)	0.53(0.0002)	4.54	<0.001*
Task: Drag-&-Drop				
Damped frequency (ωx)	0.13(0.002)	0.13(0.001)	1.62	0.56
Damped ratio (ζx)	0.53(0.0004)	0.53(0.0003)	1.68	0.05
Task: Steering				
Damped frequency (ωx)	0.12(0.002)	0.13(0.003)	2.4	0.01*
Damped ratio (ζx)	0.53(0.001)	0.53(0.001)	2.55	0.007*

Table 3.1: t-tests for the x-axis damping ratio, damping frequency and completion time for each of the mouse tasks (Clicking, Drag-and-Drop, and Steering). *Statistically significant (p<=0.01) after Bonferroni correction.

To calibrate our findings, we contrasted them with HRV. We calculated HRV out of a raw Electrocardiogram (ECG) signal. We curated the signal as recommended by Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology [137]. We found significant differences between the stressor and calm intervention phases for several metrics: mean peak-to-peak R-R ratio ($t(48) = 10.95$, $p<0.001$), Low Frequency (<0.15Hz) energy (LF) ($t(48) = -2.03$, $p=0.02$) and High Frequency (>0.15Hz) energy (HF) ($t(48) = 2.05$, $p=0.02$). We did not observe any differences between the post_stress and post_calm mouse tasks. However we did observe a significant difference for subjective ratings (0 = calm → 10 = stress): post_stress ($M = 3.9$, $SE = .25$) and post_calm ($M = 2.67$, $SE = .26$) ($t(48) = 5.86$, $p<0.001$).

To complement our work, we went on to try to find a way to automatically classify the observed differences between stress and calm. As an example, we show the damping ratio aggregated over N=52 (26 females and 27 males) performing all different motion tasks with 5 randomized repetitions each for the Clicking condition (see Figure 3.1). As it can be observed, the difference between signals is not huge, but is completely discernable.

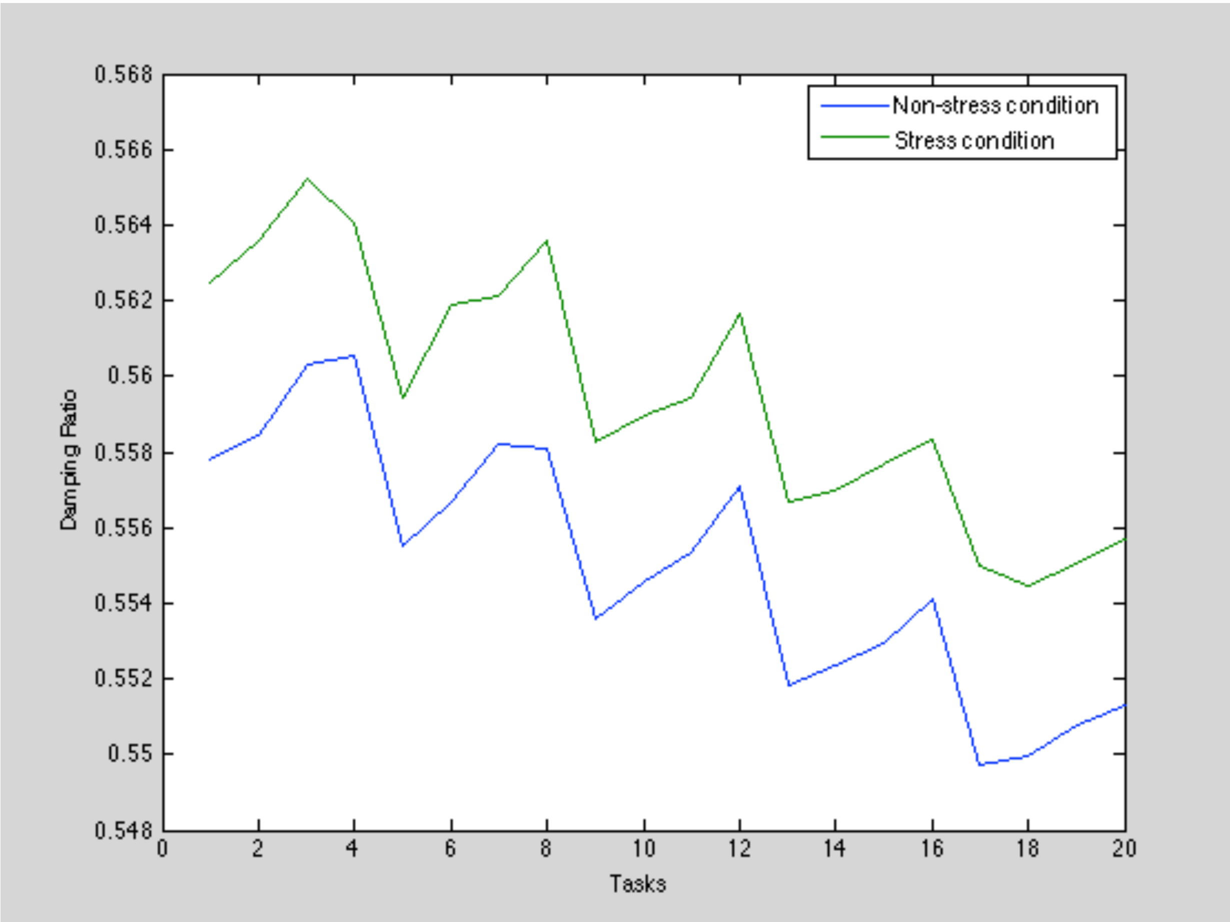


Figure 3.1: Mouse displacement damping ratio for 20 different motion tasks for mental Calm and Stress conditions [22]

For the accuracy of within-subjects stress classification, i.e. given some labeled samples for one subject as training data, we studied the accuracy of stress classification on some unseen samples. We did this by taking a random sample of k of the data points derived during the study, training a classifier on the remaining n−k points, and using this to classify the initial k samples. We used a simple model-based classifier, relying on the structure that is evident in Figure 3.1. The model has a staircase structure, i.e. we model canonical stress behavior as having a simple step-wise dependence on target distance, and a separate (step slope) dependence on target size. An advantage of this model is that it requires only knowledge of the distance of a mouse motion, not the target size. Thus an underlying logger which is not aware task or application could be used. Model accuracy peaked at about 71%

accuracy with 30 samples. As the number of measurement samples increases beyond 30, accuracy starts to fall because there are not enough remaining samples (100-k) to build an accurate model. Thus, about 10 mouse movements of a stable stress state should yield around 70% accuracy.

3.3.2 Mouse Pressure[2]

The capacitive mouse used in our work [55] is the Touch Mouse from Microsoft, based on the Cap Mouse described in [146]. This mouse has a grid of 13x15 capacitive pixels with values that range from 0 (no capacitance) to 15 (maximum capacitance). Higher capacitive readings while handling the mouse are usually associated with an increase of hand contact with the surface of the mouse. Taking a similar approach to the one described by Gao et al. [47], we estimated the pressure on the mouse from the capacitive readings.

The experiment was based on a simplified version of the Fitts' law task [80], participants were challenged to click on horizontal bars that alternatively appeared on the either side of the display. In particular, there were three different distances (200, 350 and 500 pixels) between the bars and three different bar widths (50, 85 and 120 pixels) that were randomly combined. For each of the combinations, the participant had to perform 10 repetitions. Therefore, for each task the participant had to click 90 times on the bars (3 distances x 3 widths x 10 repetitions). In order to induce a relaxed or stressed emotional state, this task was performed right after both the relaxed and stressed conditions of the expressive writing or text transcription tasks. Stress and affect was measured using three Likert scales (Figure 3.2).

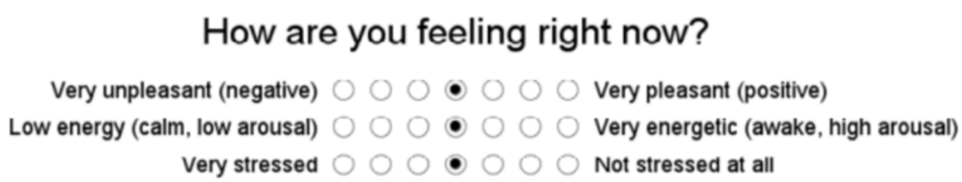


Figure 3.2: Affective and Stress Measurements using 7-point Likert scales.

To determine if people under stress handle the mouse differently, participants performed a simplified version of the Fitt's law task [80], in which they needed to click on several vertical pairs of bars of varying widths and distances from each other. Unlike the keyboard tasks, the stressor took place before the task,

---

[2] Data presented in this section is taken from: Hernandez, J., Paredes, P., Roseway, A., & Czerwinski, M. (2014, April). Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 51-60). ACM.

with either the expressive writing (for participants 1 to 12) or the transcription task (for participants 13 to 24). In this work we estimate the amount of pressure with the mouse by analyzing capacitance readings. From the raw capacitance readings of the two conditions we computed the average of all the 13x15 capacitive pixels at any point in time and created a time series for each condition (bottom-left). Finally, we estimated the overall pressure by computing the average of each series (bottom-right). As it can be seen from the capacitive readings of this example, the location of each finger can be easily identified. While the participant used 4 fingers during the relaxed condition (blue rectangle), s/he showed more contact of the pinky finger during the stressed condition (dashed-red rectangle). 18 participants showed increased mouse contact during the stressed condition, and 6 participants showed reduced contact during the same condition. The differences between the two conditions were significant for all the participants (W, $p<0.05$). 75% of the participants showed increased contact with the mouse under stress. A summary of the results for pressure difference between stress and calm could be observed in Figure 3.3.

3.3.3 Keyboard[3]

Monitoring the dynamics of keyboard usage has been widely studied in different areas such as biometric authentication [9, 96] and personality characterization [69]. Some of the main keyboard dynamics are based on latencies of the keystrokes, such as time between keystrokes or the length of time that each keystroke is pressed. One of the interesting findings when analyzing keyboard dynamics such as these reveals that the typing patterns of the same individuals vary over time and are affected by other factors such as stress or gradual changes in cognitive or physical function [96]. Thus, keyboard dynamics can provide relevant behavioral information about the affective and cognitive state of the user. Motivated by this finding, Vizer et al. [147] created a system that measured keystroke and linguistic features of 24 computer users, and were able to recognize cognitive and physical stress with accuracies of 75% and 62.5%, respectively. In a separate study, Khanna and Sasikumar [70] also used keyboard dynamics to differentiate between neutral/positive and negative emotions of 21 participants in a laboratory study. One of their main findings was that the negative emotional state was associated with more typing mistakes and slower speeds in comparison with the more neutral affective condition.

---

[3] Data presented in this section is taken from: Hernandez, J., Paredes, P., Roseway, A., & Czerwinski, M. (2014, April). Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 51-60). ACM.

Affective inference from keyboard activity was recently described by Epp, et. al. [37]. The developers used timings from individual key and short key-sequence (bi- and tri-gram) features. Data were collected naturalistically, i.e. the system monitored subjects' everyday computer use, and ESM (Experience Sampling) was used to gather self-assessments of emotional state. The system showed promising accuracy (70%-88%) for most emotion labels. While these results are encouraging, as the authors acknowledged, these raw classification rates for skewed categories masked a very modest gain over baseline classification (e.g. for excitement). Still the approach shows the power of this feature set.

We experimented with a pressure-sensitive keyboard, as described by Dietz et al. [31]. For each keystroke, the keyboard provides readings from 0 (no contact) to 254 (maximum pressure). We implemented a custom-made keyboard logger in C++ that allowed us to gather the pressure readings at a sampling rate of 50 Hz. We performed two tasks: text transcription and expressive writing. In the former stress was induced by altering the size of the characters, doubling the speed of the blink of the cursor, adding a timer and finally adding ambient noise. In the latter people were asked to write either a calm or a stressful memory for about 5 minutes. Self reported stress was significantly higher for the stressful tasks. Contrasting Electro-Dermal Activity (EDA) measured through an Affective Q-sensor bracelet did not show any differences.

Figure 3.3 (top) shows the individual differences between the stressed and relaxed distribution averages of the transcription task.

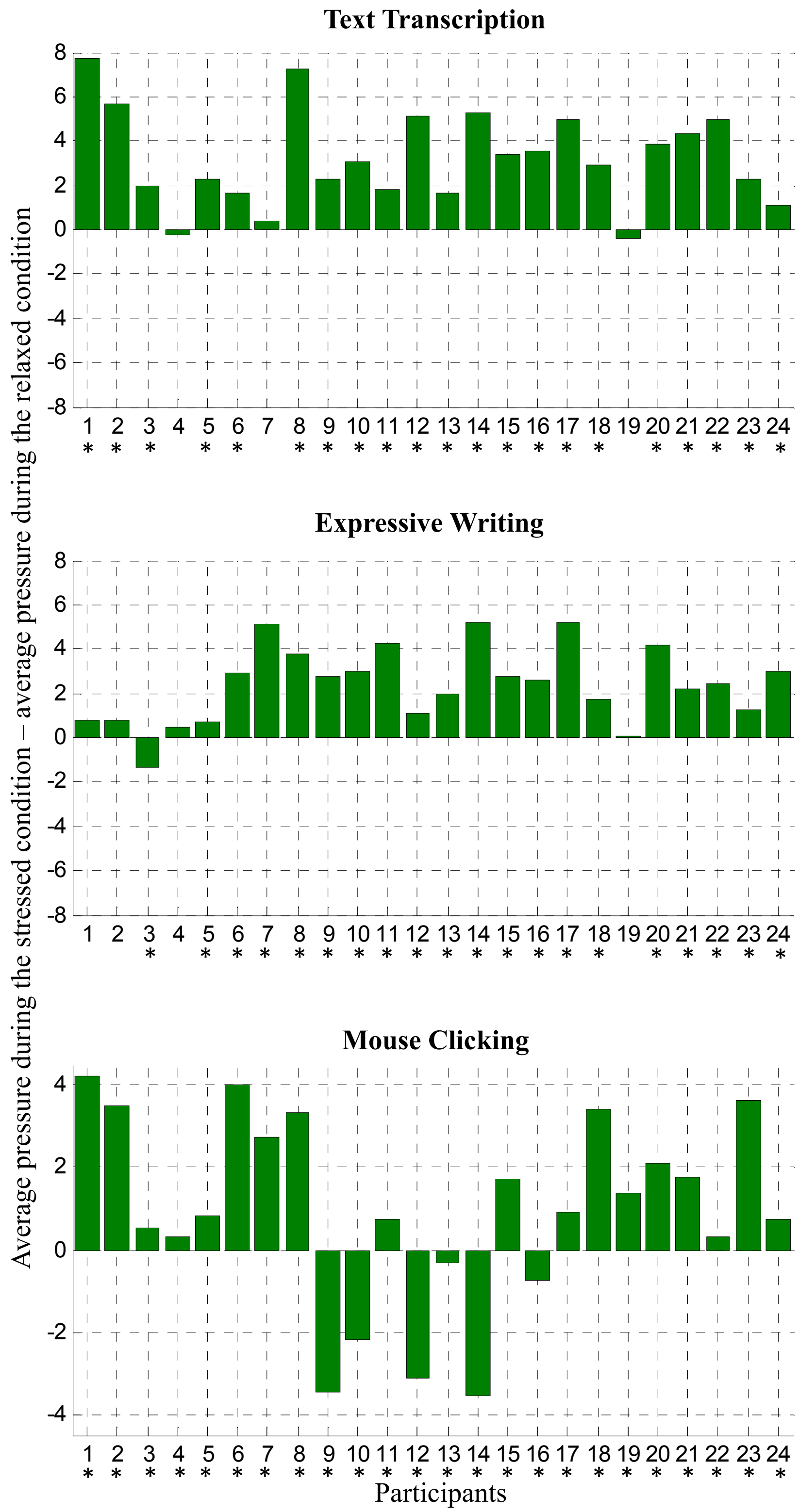


Figure 3.3: Individual differences between the averaged differential (stress – calm) signal per participant. *The difference was computed from significantly different distributions (W, p<0.05)

A positive value indicates higher average pressure during the stressed condition, and a negative value indicates higher average pressure during the relaxed condition. As can be seen, 22 out of the 24 participants (91.67%) showed higher average pressure metrics during the stressed conditions. When comparing the distributions of keystroke pressure across the two conditions, all participants except for three of them (participants 4, 7 and 19) showed significantly more pressure in the stressed conditions. These differences are similar to the ones observed during the expressive writing task shown in the middle graph of the figure. For this task, 23 out of the 24 participants (95.83%) showed increased average typing pressure during the stressed conditions, for which all participants except four (participants 1, 2, 4 and 19) showed a significant difference.

Note, however, that participant 3 showed significantly less pressure during the stressed condition. When describing a stressful memory, this participant described past episodes of depression which may have caused the decrease in keyboard pressure. Finally, the overall average typing pressure observed during the transcription task was higher than that during the expressive writing task, which is consistent with the self- reports of stress. No significant differences were found between the stressed and relaxed conditions in terms of amount of introduced characters, task duration or typing speed (amount of interactions). Furthermore, there were not significant differences in terms of pressure between the two genders. However, when comparing the average pressure values for the different task orderings in the expressive writing task, the groups were significantly different (K,H(3)=8.873, $p = 0.031$). In particular, participants that started the experiment writing about a relaxing memory (participants 1 to 6) showed smaller differences between the two conditions. This group of participants also showed lower average pressure values for the two conditions, although not significant (W, $Z = -1.5$, $p = 0.134$), than participants with other task orderings. The tendency towards decreased pressure values indicates that writing about a relaxing memory at the beginning of the experiment may positively influence the rest of the session. Collecting data from more participants and increasing the duration of the calming clip could provide additional insight as to how to prevent this effect in future experiments. Despite this one ordering effect, the differences between the two conditions were consistent across all of the other task orderings, indicating increased typing pressure during the stressed condition.

3.4 TEXT MINING

As described in chapter 1, text mining has a strong potential to discover personal insights. Another use of our semantic searching tool is to discover labels of emotion. We can evaluate the use of the tool and its querying mechanisms as a sensor-less sensing technique. To do this accuracy studies must be performed. One initial approach could be to sense long-term processes, rather than immediate changes. We could discover life events that are associated with stress or strong emotions. Interventions that require long term change could benefit from these measurements. Furthermore, if we manage to generate a large body of data with proper markers, we could potentially explore causal inference methods, as described in the prior section. With enough available text data, we can perform causal inference in a way that allows us to determine the directionality of a mental health outcome. For example, determining if mood changes lead to poor interpersonal skills or vice-versa could help describe better interventions.

3.5 MOVING FROM SENSING TO INTERVENTIONS

As observed, sensor-less sensing is an important enabling technology. It provides a tool to measure important mental health metrics without requiring any change in the user's behavior. However, to ultimately rely into a wide adoption of these techniques, it is important to understand other adoption and engagement factors such as usability, context, field use and causal inference.

3.5.1 Usability Studies

Given the new uses derived from extracting personal, and in many cases intimate, information, our research approach contemplates a fundamental focus on usability. In the case of sensing personal mental data, it is important to understand that many users may not have a clear conceptual model of this novel sensing technology. Additionally, their level of motivation and their level of engagement will be diverse and certainly should change over time.

It is relevant to examine the interaction between the user and a device that "reads" your mind and how this can generate further confusion, anxiety, or behavioral changes, as well as desire or rejection of the system. Usability studies should help understand the differences between a) treating the device as reader of one's (mental) self or b) as a companion ("pet") that is affected or altered by our mental states. Furthermore, single mode and multi-modal lab and field studies should be performed to observe adoption and engagement patterns.

3.5.2 Stress and Emotion from Sensor and Contextual Data

We will begin with lab studies with subject under induced stress and emotions. We have run some of these studies in our lab to date, using a variety of calibration methods [112] [135]. In reality it is difficult to train a stress or emotion sensor because there is no objective ground truth. While Heart Rater Variability (HRV) responds strongly to Autonomous Nervous System (ANS) tone, there are several confounds. Even chemical tests (e.g. cortisol samples from saliva) are confounded by body chemistry (esp. medications), and exhibit significant lag (minutes or tens of minutes). We make use also of induced stress, but different subjects respond very differently to particular stressors, so the best we can hope for is an increase in stress on average. This is still enough to train a model, and once trained we can study the accuracy of this model with or without the inclusion of particular sensors to measure their individual predictive value. Data for stress modeling will include continuous variables (estimates from the biometric sensors including voice, keyboard, mouse, HRV), a periodic time variable, and discrete variables (location, id of a nearby person). We will start with simple models, namely linear regression of sensor data on a consensus of “control” signals (cortisol, self-report and Tricorder HRV). Discrete changes will be modeled with additive linear coefficients. Going beyond pure supervised regression; we will experiment with latent factor models, which may expose useful patterns of thought and/or behavior, which predict stress. We have previous experience with rich factor models for activity classification from desktop activity data [121].

3.5.3 Integrated Model and Data Acquisition

Given the models as trained above the final boundary if to perform field modeling. Using the models above, it is possible to compute real-time estimates of emotional states and stress for users working at a desktop or using a mobile device. When significant changes occur in modeled emotional and/or stress level occur, the system will prompt the user to give a self-report. These self-reports will then be used for further training and model refinement. By issuing requests at times of change, we will be able to gather emotional and stress label data that should be as close as possible in time to triggers. Thus it should be of maximum value for building models of the complex dependencies on discrete emotional triggers and stressors.

3.5.4 Causal Inference

With enough available text data, we can perform causal inference in a way that allows us to determine the directionality of a mental health outcome. For example, determining if mood changes lead to poor interpersonal skills or vice-versa could help describe better interventions. Based on our text labels we could mark different observational variables, such as life events, moods, sentiment, and other variables that we could later analyze. It is important to note that causal relationships are not sufficient to define good interventions.

Another key element is to know they types of interventions, when interventions should be applied and the appropriate engagement strategies. Some of these topics will be treated in Part 2 of this thesis.

3.6 ADAPTIVE INTERACTIONS AND INTERVENTIONS

3.6.1 Feedback and Interface Adaptation

The first and most basic interaction/intervention will be to present direct feedback of different emotional states to users. Direct feedback should raise awareness and drive behavior change. Another option is the automatic alteration of the interface to better adapt to the emotional state of the user. The interface could be adapted either at the background or foreground level [62] to help the user either maintain attention in the task at hand during a stressful, but productive state, or change and relax during a frustrating or overwhelming episode. Interface adaptation to human affect has been used in airplanes, using previous knowledge, self-reports, diagnostic tasks and physiological sensing and changing the interface at the content and format levels [60]. Work has also been done to adapt computing interfaces to cognitive and affective changes (the latter captured through changes in facial expressions) [34]; however little work has been done to unobtrusively sense affect and actively adapt computer interfaces to affective changes.

We plan to explore the use of sensor-less sensing to monitor different emotions commonly present during interactions with computing devices to help people use them more towards the accomplishment of their goals and to optimize the emotional engagement associated with it. In terms of mental health, the ability to maintain an "appropriate" level of stress, needed to fulfill the task will be beneficial especially in productive settings like the office or school.

3.6.2 Emotional Regulation and Psycho-education Interventions

Emotional regulation interventions such as calming technologies present a good opportunity to help deliver users that present initial (moderate) symptoms of stress or other emotional changes with a brief and effective intervention that would help people avoid unnecessary emotional alteration. Paredes, et. al. [112] present an example of some mobile individual and social interventions that are brief and usable without the need of additional hardware and leveraging intrinsic and social aspects. Coyle, D. et. al. [30] has explored gaming interventions focused mental health based on Cognitive Behavioral Therapy (CBT). Chapter 2 describes some social and gaming concepts that could benefit from a sensor-less sensing technology. More specifically, section 2.6 describes the machinima (machine + cinema) technology for the creation of short movies that can deliver some life skill or CBT-based stress reduction. Section 2.6.3 describes how machinima can be built from a CBT manual [101].

3.6.3 Foreground and Background Interventions

Additionally, novel designs of foreground and background interventions delivered via the graphical unit interface (GUI) could be part of operating systems or applications. Some ideas in that could be developed are: a. the use of desktop themes, screen savers, menu bars, etc. could be used to deliver soothing messages or color combinations that can help reduce stress, b. the modification of fonts or illumination of the screen to help improve reading during moments of emotional arousal, c. help block or reduce the number of cues presented through automated messages, such as incoming email or chat, to help maintain focus during highly stressful situations.

3.7 CONCLUSION

As described in this chapter, sensor-less sensing opens novel possibilities to work with a new suite of psychophysiology signals. It allows the use of movement and the Somatic Nervous System (SoNS) as a source of reliable data through movement and pressure. It is exciting to observe that we can take advantage of widely used devices, such as mice and keyboards. We also discuss the repurpose of our semantic search technology described in chapter 1 as a key enabler of using text as a sensing stream.

By sensing without disrupting the user's workflow, we considerable reduce barriers for adoption. However, we discuss also the importance to not only rely on sensing as an enabler of change. Additional research is fundamental to be able to move from sensing to an efficacious intervention deployment. We discuss factors such as usability, context and field studies as a key complement to any sensing approach. Finally, we discuss the novel types of interventions that would be much harder to implement without the ability to use unobtrusive sensing techniques. Part 2 presents a body of research that takes advantage of some of our conceptualization and sensing innovations to generate novel "opportunistic" interventions.

Part 2
Opportunistic Interventions

"Opportunistic Interventions" repurposes well-adopted devices and media into effective interventions. We aim at increasing efficiency (adoption and reduced attrition) beyond efficacy. We discuss three important types of interventions: Suites, Multi-modal, and Environmental. Key elements that transcend these types are:

> *Novelty*. An efficacious intervention can become obsolete due to boredom. Humans adapt to its effects, and their interest decreases with time. Suites of interventions have the potential to deliver new content over time. Fresh interventions and authoring systems can reduce novelty effects.
>
> *Context matters*. An intervention can become a stressor under the wrong context. A person receiving a personal message could interpret it as uplifting or as embarrassing. Context data is important and should be input to a successful intervention approach.
>
> *Introductory period*. Users with no previous training can find an intervention taxing. For example, people with no experience in deep breathing may hyperventilate. A training period could improve efficacy and reduce attrition.
>
> *Negative outcomes*. Interventions can have negative affective outcomes if poorly designed or implemented. Factorial experimental design can help flush out undesired outcomes. Multi-modal approaches should take into account potential interaction effects. Environmental interventions should consider prior perceptions.

Chapter 4 studies how to design a suite of mobile and wearable interventions. Our focus is in diversity while maintaining efficacy. Chapter 5 presents interventions harvested from the web. Popular apps become efficacious interventions. Machine learning policies use contextual cues to recommend the best intervention. Chapter 6 discusses the difficulties of multi-modal interactions in wearable systems. Haptic and sound interaction effects are analyzed. Finally, Chapter 7 showcases novel urban LED lights that elicit emotional outcomes in pedestrians.

Chapter 4
Repurposing Mobile and Wearable Systems

This chapter describes design explorations for stress mitigation on mobile devices based on three types of interventions: haptic feedback, games and social networks. This chapter offers a qualitative assessment of the usability of these three types of interventions together with an initial analysis of their potential efficacy. Social networking and games show great potential for stress relief. Lastly, this chapter introduces key findings and considerations for long-term studies of stress mitigation in HCI, as well as a list of aspects to be considered when designing calming interventions.

4.1 INTRODUCTION

Stress is an exacerbating factor for many physiological and psychological illnesses [26]. Three promising avenues can mitigate stress. First, the sense of touch has been regarded as an important communicator of empathy and calmness [126]. Second, social support is a valid tool to reduce stress [54]. Third, playing games is considered a distraction that could be used as a way to relieve stress [67]. We want to investigate novel HCI approaches to calm people down in the early stages of the accumulation of stress. We want to answer two questions: (1) Is it possible to reduce stress through interactive techniques? (2) Which modalities or interfaces are most promising to calm people down? Specifically, we investigate three different approaches suggested by prior work [12, 54]: haptic feedback, where vibrating motors stimulate acupressure points; interactive games, where game play can reduce stress; and social network interactions.

4.2 BACKGROUND AND RELATED WORK

Appropriate feedback is crucial to behavioral change. Positive psychology [131] is currently emerging as a new way of inducing behavioral changes including helping people calm down with appealing cues. Evidence Based Therapies administered via the Internet [102] have been showing promising results. As an example, Cognitive Behavioral Therapy (CBT) [12, 115], which uses several techniques for habit change teaches people how to recognize their sources of stress and block the negative associated reaction [8]. Narrative Therapy focuses on constructing conversations to help people be satisfied with their state of being [151].

Recent studies have demonstrated the value of haptics as a therapeutic tool [144], where different haptic techniques are used to support different mental health therapies. The game —Relax to win implements bio- feedback (controlling the game through bio signals) as a mechanism to help people relax [122]. In addition, several game mobile games and web applications have recently appeared that are designed to calm players.

4.3 IMPLEMENTATION

4.3.1 Prototypes

We have built four prototypes to explore the usability and efficacy of mobile interventions to manage stress:

Social Networks

We built a text-based interface using SMS to deliver alert messages (see

Figure 4.1). We worked with participants and a small circle of their friends to test responsiveness of this method. This setup was chosen because of the power of intimate communication between friends and family; and because of the pervasiveness of SMS messages.

Figure 4.1: An SMS message used to prompt the closest contacts in the user's (inner) social network.

Playing Games

We used commercially available mobile games with simple tasks, such as mazes and basic interaction games (tilting, moving, rotating). We allowed participants to choose games at their own discretion. This intervention was chosen because of the distraction it provides. Figure 4.2 shows an off-the-self maze game.

Figure 4.2. One of the games we used involved maneuvering a ball in a maze

Guided Breathing

We employed vibrotactile feedback to guide breathing patterns. Figure 4.3 shows a bracelet with two eccentric rotating mass (ERM) vibration motors. We coached participants to breathe according to well-known deep- breathing techniques, where a haptic stimulus indicates the breathing rhythm. Correct breathing rhythm is one of the key elements to achieve a soothing effect.

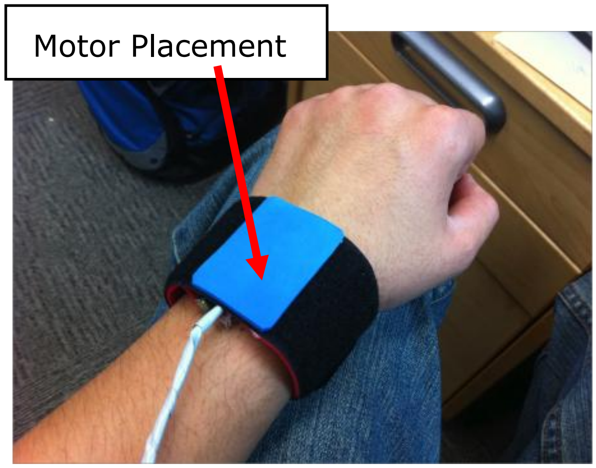


Figure 4.3: Vibrotactile bracelet with two eccentric rotating mass (ERM) motors for guided breathing and acupressure stimulation.

Guided Acupressure

We employed the aforementioned bracelet to stimulate acupressure points in the wrists and the chest (when the wrist is held to the sternum); these points are known to reduce stress (Figure 4.4). We employed a Wizard of Oz technique to control the timely application of this stimulus to the participants.

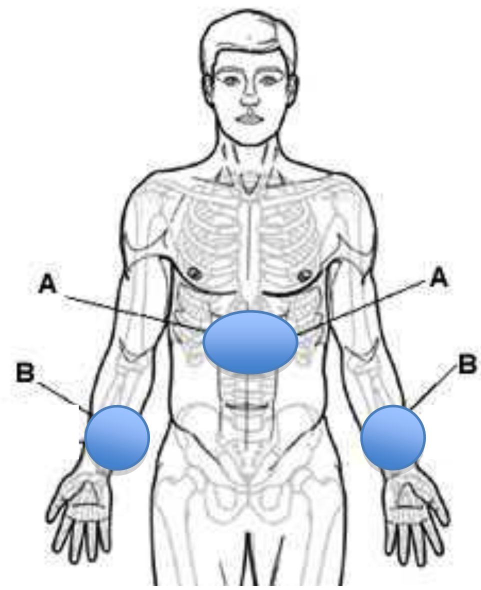


Figure 4.4: Accupressure points for relaxation.

4.3.2 Common use scenario

Joe, a college student, has been preparing for a presentation for his class. He is nervous about public speaking. He starts to panic and worries that he will make a fool of himself! Joe's mobile device, connected to a Berkeley Tricorder sensor [106], detects that he is stressed and activates his haptic-guided breathing bracelet. It also sends an SMS alert to his close friend Ben. After a few moments of deep, guided breathing, Joe receives an SMS text from Ben with a funny joke and a line of encouragement. Joe calms down and is now ready for his presentation.

4.4 STUDY DESIGN

The focus of this study is to evaluate promising calming technologies. We first investigated the efficacy and potential use of such technologies in a lab experiment. To gather data on the effectiveness of our interventions we collected both used objective (biometric), and subjective (psychometric and self-report) data.

4.4.1 Objective Data

Electro-Dermal Activity (EDA) and Electrocardiogram data (ECG) are known indicators of stress, if the subject is not engaged in strenuous physical activity. We gathered this data using an ambulatory sensing device, the Berkeley Tricorder. We used four features:

1. *ECG Heart Rate (HR) (positively correlated with stress)*

2. *ECG Heart Rate Variability (HRV) (negatively correlated)*
3. *Electro-dermal conductance (EDC) (negatively correlated)*
4. *EDA Variability (negatively correlated)*

These biometric data are valuable because they can assist in verifying the subjective stress assessments of

users. Awareness of personal stress levels varies between people; this limits the validity of self-reports.

4.4.2 Subjective Data

We applied three commonly used scales to measure perceived stress:

1. *State-Trait Anxiety Inventory scale (STAI)*: analyzes general and momentary feeling of anxiety
2. *Subjective* assessment of stress: 0 – 10 Likert scale
3. *Life Events Questionnaire*: evaluates long-term stress accumulation via questions about events such as the death of a relative, divorce, job loss.

Stress levels were induced via two mental stressors:

a. *Stroop Color Word Test*: Presents a series of words that describe colors, but using a different color of ink at increasing speeds.
b. *Recurrent Subtraction Math Test*: Which includes penalties for mistakes and for long response times—if users made a mistake or took too long, they had to start over.

4.5 EVALUATION

4.5.1 Baseline

We recruited 20 participants (13 male and 7 female) among students in our institution. The evaluation consisted of two phases, each one with 6 identified stages: (1) Arrival, (2) Calming, (3) Stroop Test, (4) Wait/Anticipation, (5) Math Test, (6) Calming (Figure 4.5). The first stage is used to gather psychometric and subjective data at the moment of arrival. During the calming stages, participants completed positive thinking and visualization exercises and we incentivized them to speak descriptively about beautiful and soothing situations. No other interventions took place.

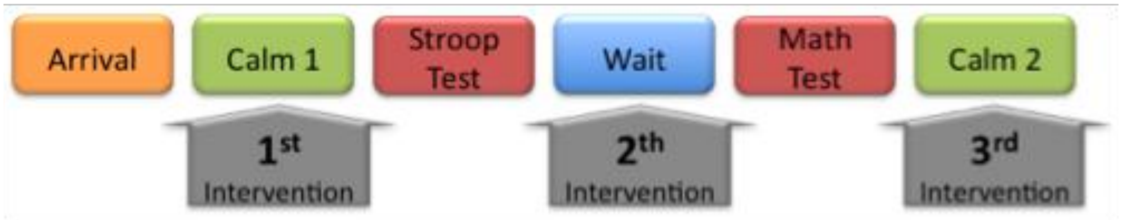


Figure 4.5: Experiment Stages

4.5.2 Randomized Experiment

In the second part of the experiment, administered two weeks later, participants completed the same stages - but this time we applied the four interventions (Social Network, Playing Games, Guided Acupressure and Guided Breathing) during stages 2, 4 and 6, in randomized order, in order to measure their efficacy to relief stress. Participants were assigned in random order to each of the potential combination of 3 interventions to obtain a within-subjects comparison of the interventions. We expected to obtain results that were either better or similar to the visualization and positive thinking stages from part 1. Finally, at the end of part 2 we gathered closed-form (Likert scales) qualitative information about likeability, potential benefit, and perceived efficacy of the interventions. We also gathered open-ended information about improvements as well as suggestions for new interventions.

4.6 RESULTS

As seen in Figure 4.6, the subjective scale expressed the expected pattern of stress and calmness in different stages. At an aggregate level, stressors did raise perceived stress levels and the calming stages did manage to lower the stress level. Participants' subjective stress ratings were ranked to account for differences in individual rating scales. Ratings for each stage were ranked from 1 to 6: the stage with the lowest subjective stress rating was ranked as 1; the highest stress rating was ranked as 6. Using a multiple comparisons Friedman test, the variation between stages was found to be statistically significant ($p<0.001$). Additionally, in part 2, the subjects left the test with a lower level of stress than the level they entered with ($p<0.001$). No statistical significance was found regarding the effect of each intervention. In a paired t-test comparison between parts 1 and 2, no significant effect was observed between interventions. This indicates that our interventions had similar effects to the positive thinking and visualization techniques of part 1.

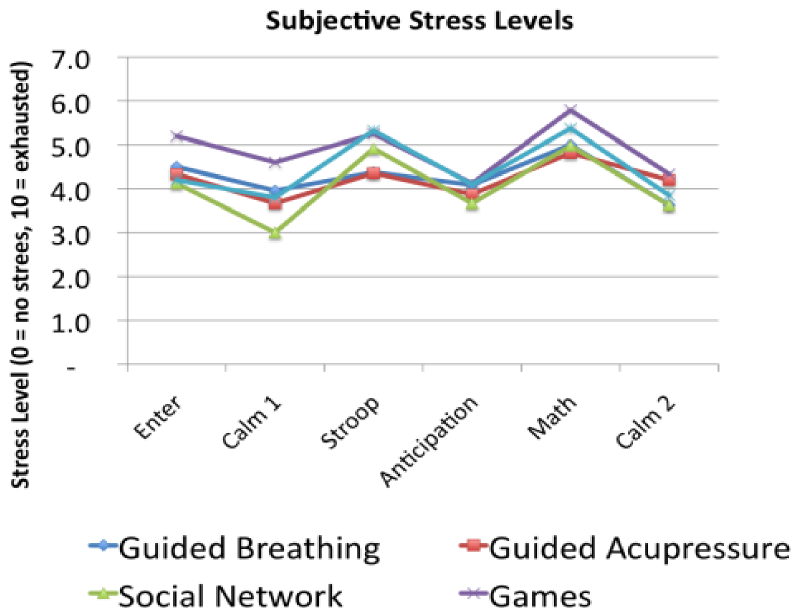


Figure 4.6: Subjective stress levels for each intervention across the different stages on part 2.

Figure 4.7 presents the normalized data gathered from the usability questionnaire. The social network support and breathing interventions have stronger, more uniform support. Acupressure and games have some strengths and noticeable weaknesses.

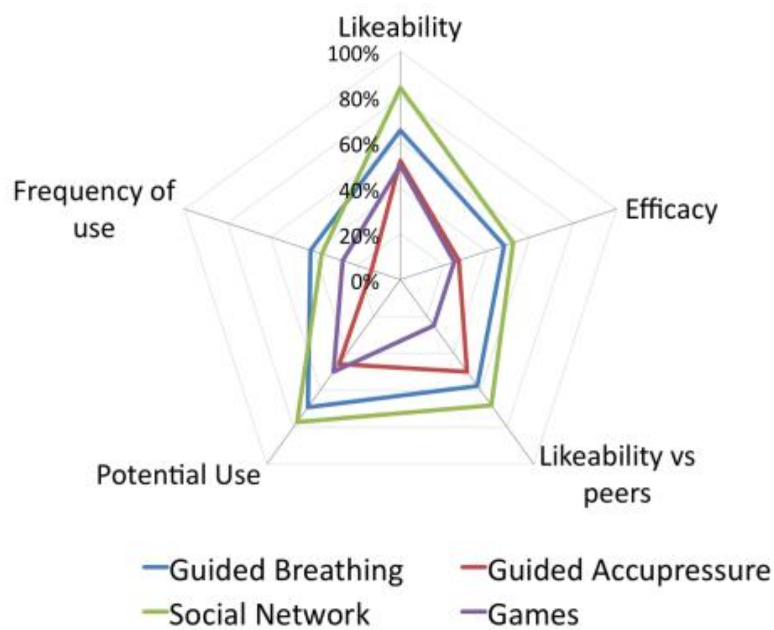


Figure 4.7: Comparison of the four interventions. Social Networking and Guided Breathing had the most uniform

A preliminary Principal Component Analysis (PCA) on the objective data showed that all the aforementioned biometrics used to evaluate stress provide 86% correlation with the expected results. This result indicates that these biometrics are relevant to the future problem of inferring stress from ECG and EDA samples, in conjunction with subjective data. Further analysis is important to reduce the set of components. Additionally, we performed a pairwise t-test of the aggregated values for the stress stages (Stroop and Math) and the calm stages (Calm 1 and Calm 2). We observed that the difference was statistically significant ($p<0.001$), which means that indeed, EDA follow the subjective stress states at an aggregate level. However, at an individual level there are many differences and gaps, which suggests that careful analysis and further design may be needed to guarantee usability. Figure 4.8 shows an example of EDA signal levels plotted against samples, where oscillations show the changes between stages, inversely comparable with the peaks observed in Figure 4.6. Lower values are correlated with high stress stages, while higher peaks with calm ones.

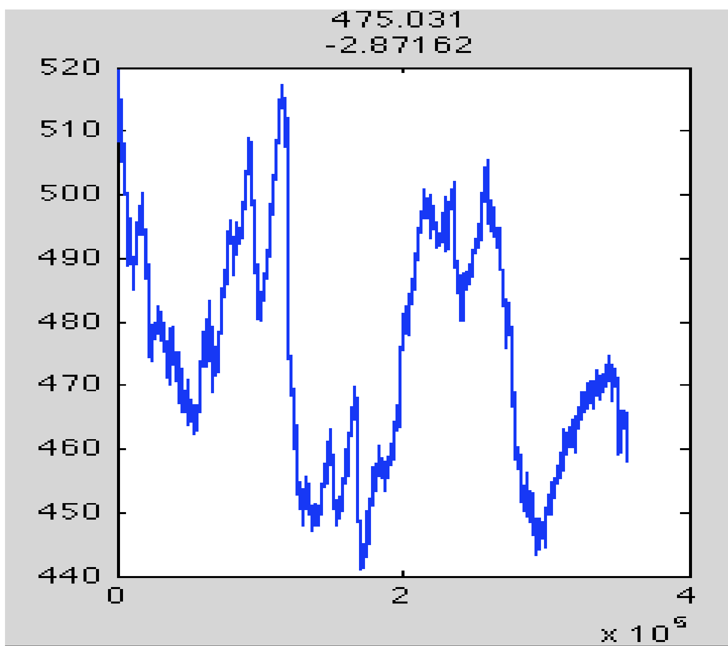


Figure 4.8. Electro-Dermal Activity (EDA) oscillating around 540 mV

Figure 4.9 shows a summary of the levels of the different biometrics. On an aggregate level many expected behaviors were observed: lower EDC mean, positive EDC change and lower HR for the calm stages and their inverses for the stress stages. HRV did not show the expected behavior, showing a higher value for stress.

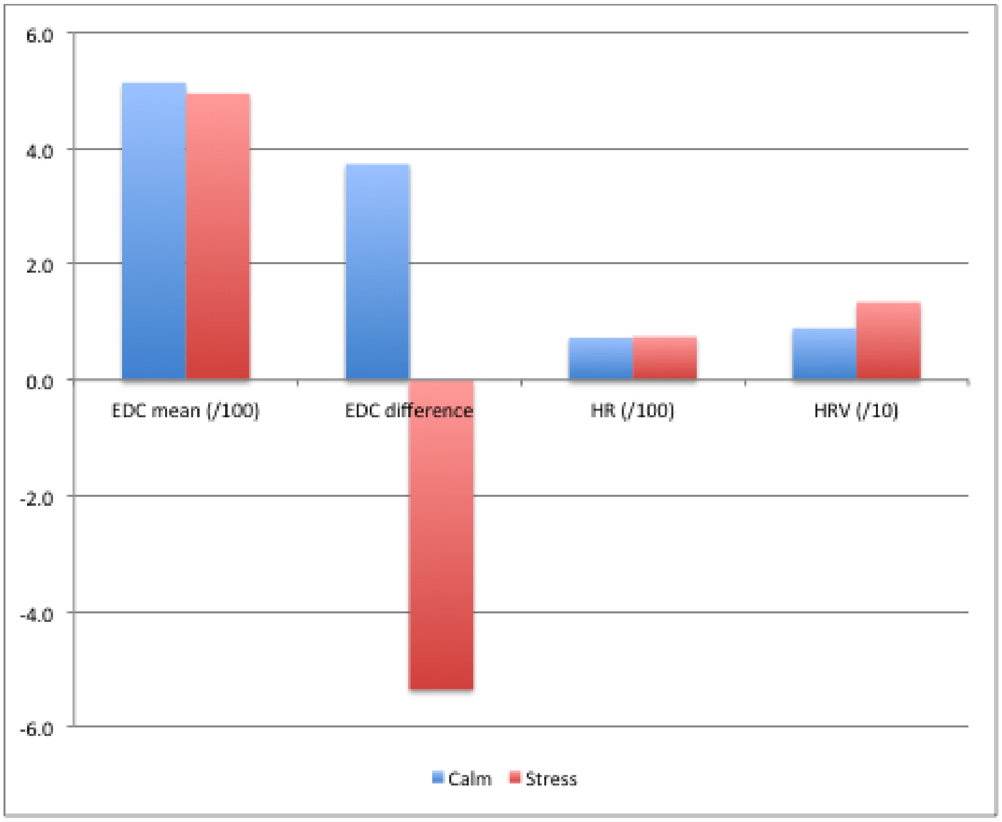


Figure 4.9: Biometric signals comparison between Calm and Stress

4.7 IMPLICATIONS FOR DESIGN

4.7.1 Suite of Interventions and Context

Many interventions could coexist in a system, and its application should be based on four factors, all in relationship with context:

Volatility: assign an intervention to a situation that carries the lowest risk of becoming a stressor. As an example, breathing techniques may not be appropriate in a humid or hot climate.

Interruption time: assign an intervention that has the adequate amount of interruption time. As an example, long interventions may not be appropriate during a meeting break.

Media: assign an intervention appropriate to the time and place. As an example, sound-emitting interventions may not be appropriate in an office.

Habituation: all the interventions run the risk of growing old. Having a suite of interventions that changes over time, even for similar contexts, will be necessary to maintain the users' interest.

4.7.2 Design improvements

Some improvements to the different interventions have been identified based on the qualitative data:

Social Networks:

Timely response: by increasing the number of contacts or by maintaining a message repository to be used when no contacts are available.

Help button: Allow the person to request help from their contacts.

Breathing:

Pressure feedback: may resemble the act of breathing better than vibration.

Training and adaptation: Gradually increasing speed and frequency of breathing could be useful especially for people not familiar with deep-breathing techniques.

Acupressure:

Training and adaptation: Some participants did not manage to wear the bracelet correctly, and others found it to be too ―novel‖. Some users mentioned that as time passed they felt more at ease.

Other acupressure points: Some users mentioned their interest to apply the device in other points such as neck, arms and legs.

Playing games:

Personalization: a personalized suite of games will be important to choose the best-suited games to calm people down.

Passive games: some participants mentioned that active games gets them stressed and/or ―hooked‖. Games with very simple tasks could be more beneficial. Games deserve further study to define the right characteristics to calm people down.

4.7.3 Long-term (real-life) usability

As described by Muñoz [100] to achieve long-term usability, self-help interventions should have: a *rational* (mental model), *education/training* phase, *guaranteed usage* in the real world and *attribution* of benefits to the tool. In the case of the different tools mentioned in this study we can obtain improvements in all these areas, however longitudinal and potentially ethnographic studies may be necessary to reach appropriate conclusions. Additionally, a longitudinal study will help gather real- life data to improve the way biometrics are used to infer stress levels, specially in situations where stress is either necessary to function, or in situations where patterns generated from normal physical or mental activity could be confused with stress patterns.

4.8 CONCLUSION

We are currently analyzing the biometric data to further add value to the selection of the most promising techniques. These techniques will be used in a larger experiment where biometric, subjective and psychometric data will be used to infer ambulatory stress and analyze potential interventions. Our study provided some design guidelines, as well as a perspective on promising interventions. Some of the key findings are:

Potential efficacy of mobile interventions: There is promise that mobile interventions can potentially calm people down in the earlier stages of stress.
No significant difference between interventions: Further study is needed to find differences and/or to optimize intervention designs.
Social networks leverage humor and intimacy: Intragroup virtual interaction can be leveraged to reduce individual perceived stress levels.
High volatility: All interventions showed a degree of volatility (risk to become stressors) with context.
Gaps between subjective and objective stress data: Further study of the discrepancies between subjective and biometric data could provide important information to improve the way mental states are inferred.

Chapter 5
Harvesting Web Interventions

Stress is considered to be a modern day “global epidemic"; so given the widespread nature of this problem, it would be beneficial if solutions that help people to learn how to cope better with stress were scalable beyond what individual or group therapies can provide today. Therefore, in this work, we study the potential of smart-phones as a pervasive medium to provide therapy for the general population - "popular therapy". The work melds two novel contributions: first, a micro-intervention authoring process that focuses on repurposing popular web applications as stress management interventions; and second, a machine-learning based intervention recommender system that learns how to match interventions to individuals and their temporal circumstances over time. After four weeks, participants in our user study reported higher self-awareness of stress, lower depression-related symptoms and having learned new simple ways to deal with stress. Furthermore, participants receiving the machine-learning recommendations without option to select different ones showed a tendency towards using more constructive coping behaviors.

5.1 INTRODUCTION

Stress is considered the modern-day killer epidemic [66]. Many physiological and mental disorders are associated with stress [5, 88]. Sadly, although many people (69%) recognize that stress is a big problem, only a small number (32%) actually know how to deal effectively with it [5]. Recent discoveries have shown the importance of coping with stress in a constructive way in order to reduce its damaging effects [130]. When undergoing stress, our body experiences a series of physiological changes, colloquially named the "fight or flight" response [123]. In the past, our ancestors living in the prairies relied on this mechanism to survive threats, such as being chased by an animal, or attacked by another tribe, etc. Modern humans are exposed to many psychological stressors - such as an imminent paper deadline, an interview, an important presentation, etc. However, often there is no practical way to "fight" or "flight" from these stressors. Coping constructively with such modern stressors is in many ways a skill we have to learn.

Several theories have driven the creation of effective therapeutic interventions [23, 33]. However, despite their efficacy, these interventions are not always efficient. Among several of the challenges, the interventions often suffer from two delivery problems: low adherence and low engagement rates. Research into how to improve these efficiency metrics is actively being pursued in the psychological community. In practical terms, the challenge to deliver effective interventions in real life can be summarized with the following question: how can we design the "right" intervention(s) to be delivered at the "right" time(s)?

In this chapter we focus on the first half of this challenge, i.e., "what" should a mobile app recommend when the user needs an intervention in any real life setting; leaving the "when" as future research. We conducted a study to verify three main questions:

Can we repurpose popular applications and web-sites as stress management micro-interventions?

Can the efficiency of such interventions be greatly improved by personalizing them to each individual and their context?

Can we gently move people's stress coping tendencies from destructive to constructive ones over time?

We designed a system based on an adaptive "learn-by-doing" model that allowed us to verify these claims.

In the following sections we explain the current attempts used in the HCI community to create computerized mental health interventions. We explain our novel intervention authoring process and the resulting group of micro-interventions derived from it; we present the details of the machine-learning (ML) system, its sensor inputs and algorithms to successfullymatch a user's intervention request; we describe the mobile app we designed to record context data, enable an Experience Sampling Method (ESM) and deliver interventions; and we finally present study results and their implications for design and future research for recommender systems that leverage popular media.

5.2 BACKGROUND WORK

Contemporary research of technology for mental health has been mostly focused on sensing symptoms. Movements such as the Quantified Self and Wearable Computing are driving research focused on the development of new and adequate sensors that will enable clinical research. Much less research is focused on the delivery or the enablement of new therapies. We cite below a few examples of these studies, most of them extending current therapies and others exploring novel technologies.

5.2.1 Cognitive Behavioral Therapy (CBT) based technology

The most notable and successful use of technology for mental health is Computerized-CBT (CCBT) systems. The most relevant are: MoodGYM [98] FearFighter [39] and Beating the Blues [11]. These systems provide effective treatment even covered by health insurance in some countries. Another interesting effort is the use of gaming platforms for CBT enhancement [30]. Online technology has also been utilized for CBT-based smoking cessation [100], and recently for personalized CBT treatment [33]. An example of a mobile system leveraging CBT concepts is the PTSD coach [120]. This app teaches patients with Post Traumatic Stress Disorder (PTSD) skills to manage their anxiety or depressive episodes.

5.2.2 Stress Technology Intervention R&D

Specifically focused on stress, Paredes and Chang's CalmMeNow [112] presents four mobile interventions: social messages, breathing exercises, mobile gaming and acupressure. Among the relevant findings was a confirmation that there is a fine line between an intervention being effective and actually becoming a stressor, if applied in the wrong context. MoodWings [81] explores a wearable biofeedback design that helps people become aware of their stress to help regulate it. Maybe not surprisingly, during a driving task, stress actually went up, but driving performance improved significantly. This underlines the importance of stress as a normal reaction when performing demanding tasks, and that it is not always advisable to reduce it, but to perhaps simply keep it under control. More recently, wearable devices have been developed to help regulate breathing patterns and increase breathing mindfulness [99]. It is worthwhile mentioning also the existence of various meditation and breathing commercial applications that help people learn relaxation and mindfulness.

Unlike previous studies, the goal of this work is not to find “the best” intervention. Instead, we argue that there is not a one-size- fits-all intervention. Therefore, we present methods that allow the authoring of many interventions and matching them to individuals based on their personalities and current needs.

5.3 STRESS MANAGEMENT SYSTEM DESIGN

We designed and implemented an application for Windows Phone 8.1 and cloud based services to support the delivery of micro- interventions, providing recommendations on the interventions, collecting user feedback and collecting contextual information. In the following sub-sections we provide more details about its main components.

5.3.1 Micro-Intervention Authoring System

Design Objectives

Our intervention authoring system design objectives were two: 1) maximize engagement and 2) discover online activities that could be used by the general population to reduce stress.

1. Maximize Engagement: Engagement is a key component of therapy adoption. Eysenbach’s work on attrition science [38] explains the importance of understanding the differences of intervention adoption between traditional drug trials and eHealth systems. He proposes metrics that capture not only the intrinsic efficacy of the interventions, but also its usability efficiency. Additionally, Schueller’s research on personalized behavioral technology interventions has shown the importance for patients to choose their own interventions as one way to improve engagement [131]. Finally, Doherty describes four strategies for increased engagement in online mental health: interactivity, personalization, support and social technology [33].

2. *Stress Reduction Online Activities*: Psychology research has shed some light on people’s natural abilities to deal with stressful situations during daily life. Bonanno has studied people’s innate characteristics to recover from stressful situations [14], Lazarus has described the strategies people use to cope [75] and the field of positive psychology studies the ways people use their strengths to reduce the impact of stress [132]. However, people deal on a daily basis with stress using simple physiological and psychological activities,

such as breathing before reacting negatively, giving meaning to hardship, laughing, etc.

It is reasonable then to assume that the recent trend towards mobile apps should reveal people using these apps for stress reduction activities. Some apps do share some characteristics similar to psychotherapy interventions, such as the following: keep us distracted away from our problems, record personal progress, organize thoughts, socialize, etc. Given this observation of the online world, we decided that it was worthwhile to explore the use of web apps as proxies to psychotherapy micro- interventions.

Mapping the Design Space

Our design space can be mapped as the intersection of stress management psychotherapy and popular web apps. We started by mapping the most commonly used stress management psychotherapy approaches. Then we grouped these approaches into four categories: Positive Psychology, Cognitive Behavioral, Meta-cognitive and Somatic (see Table 5.1). We chose this classification based on two premises: a) it corresponded to a *theoretical framework* as a good approximation to therapeutic approaches and was accepted by clinical psychology collaborators and b) it was simple enough to be *presented to users* using friendly nametags. As mentioned earlier, socialization is an element associated with improved engagement [33]. Therefore we further sub-divided our four intervention groups into interventions that could be performed alone (individual) or with or for others (social). In parallel, we mapped the top web apps [140] and top (Windows Phone) apps [152] and games [153]. We chose best rating metric as a proxy for popular/engaging apps, i.e., those with high levels of adoption and user satisfaction.

Micro-Intervention Structure
We wanted to design micro-interventions that followed some of the effective usability characteristics described by Olsen [111], i.e. a micro-intervention that could be designed by diverse *design populations* (i.e., psychologists, caretakers or even users), be used in *combination,* and *scaled up* easily. We boiled down the micro- intervention format to a minimal expression using only two components: a text *prompt* that tells the user what to do and a *URL* that launches the appropriate tool to execute the micro- intervention (see Table 5.1 for examples). Furthermore, we constrained the micro-interventions to be representative of one of the psychotherapy categories, and performed in a short time (approximately less than 3 minutes) to maximize usage scenarios.

Therapy Group	Therapy Techniques	Group Icons and Names	Micro-intervention Samples
Positive Psychology Focus on wellness and well-being, and making the positive aspects of life more salient.	- Three good things - Best future self - Thank you letter - Act of kindness - Strengths - Affirm values	**Food for the Soul** (Individual) **Social Souls** (Social)	• Individual: **Prompt**: *"Everyone has something they do really well... find an example on your FB timeline that showcases one of your strengths."* **+ *URL***: http://www.facebook.com/me/ • Social: ***Prompt***: *"Learn about active constructive responding and practice with one person"* **+** ***URL***: http://youtube.com/results?search_query=active+constructive
Cognitive Behavioral Observe thoughts, their triggers and their consequences, entertain alternatives, dispute them, etc.	- Cognitive reframing - Problem solving therapy - Cognitive Behavioral Therapy - Interpersonal Skills - Visualization	**Master Mind** (Individual) **Mind Meld** (Social)	• Individual: ***Prompt***: *"Challenge yourself! Replace an unpleasant thought with two pleasant ones. Write the pleasant ones down."* **+** ***URL***: http://www.shrib.com/ • Social: ***Prompt***: *""Try to think what new perspective these news bring to your life and share them with others."* **+** ***URL***: http://www.huffingtonpost.com/good-news/
Meta-cognitive Respond to ongoing experience episodes with emotions that are socially tolerable and flexible to permit spontaneous reactions or delay them as needed.	- Dialectic Behavioral Therapy - Acceptance and Commitment Therapy - Mindfulness - Emotional Regulation	**Wise Heart** (Individual) **Better Together** (Social)	• Individual: ***Prompt***: *"Shall we play a short game?"* **+** ***URL***: http://www.magicappstore.com • Social: ***Prompt***: *"Write down a stressful memory of another person, imagine it flies away and disappears, and then destroy it.* **+** ***URL***: http://privnote.com
Somatic Exercises to shift physiological signs of arousal.	- Relaxation - Sleep - Exercise - Breathing - Laughter	**Body Health** (Individual) **Social Time** (Social)	• Individual: ***Prompt***: *"Time for a quick stretch! Try some of these for a few of minutes…"* **+** ***URL***: http://m.pinterest.com/search/pins/?q=office stretch • Social: ***Prompt***: *"Cats are hilarious except when they want to eat me. Check out a few of these and show it to your friends."* **+** ***URL***: http://m.pinterest.com/search/pins/?q=funny cats

Table 5.1: Micro-interventions design matrix. Therapy groups are subdivided into individual and social groups. Each therapy group has a friendly icon and name. The micro-intervention format consists of a Prompt plus a URL.

Web apps and psychotherapy intersection

With these design elements in hand; we proceeded to brainstorm a long list of potential micro-interventions that mapped into the psychotherapy groups. This was a two-way process; we used the psychotherapy descriptions and techniques as a guide to "harvest" activities that could be applied using popular web apps (or one of their features); and vice versa, we choose some "cool" web app (or one of their features) that could be categorized into one of the psychotherapy groups. We chose two micro-interventions per group to account for a total of 16. Table 5.1 shows 8 of them.

Friendly Titles and Icons

To finalize our design process, we substituted the theoretical therapy group names with "friendly" names and icons that could be accepted by the users. We wanted to avoid names that would make people feel as if they were in therapy, and rather use names that were fun and memorable. For example, we changed Positive Psychology (individual) to "Food for the Soul" and Somatic (social) to "Social Time". See Table 5.1 for the 8 Group Names and Icons list and Figure 5.1b for a screenshot. We added a one-sentence motivational slogan per group (not shown in Table 5.1).

5.3.2 Intervention Recommender System

The goal of the recommender system was to match interventions to the personal traits of each individual and their temporal context. For example, asking someone to join you for a drink of water may be an efficient coping strategy, but one may not be able to exercise it if he or she is at home by him or herself. In order to learn the matching, we proposed modeling this problem as a contextual multi-armed bandit problem [19]. In this setting, the learning algorithm tries different interventions and learns from the feedback it gets. More specifically, we have trained a model to predict the expected stress reduction of each intervention for an individual at a given context. Based on these estimates, the recommender selects an intervention by leveraging a tradeoff between *exploiting* (refining) the best interventions and *exploring* those interventions that were not used enough to gauge their effectiveness.

Input Features

The recommender system receives both user and contextual data. *User data*, such as Personality and trait data, was obtained from a pre-study survey and self-reported mood data was obtained by implementing an Experience Sampling Method (ESM) (see next section for details). Table 5.2 shows the user's parameters that were used.

Data Type	Parameters
User Traits	- Personality: BIG5 (agreeableness, conscientiousness, extraversion, neuroticism, openness) - Affect: Positive and Negative Affect - PANAS - Depression: PHQ-9 - Coping Strategies: CSQ - Demographics: gender, age, marital status, income, education, employment, professional level - Social network usage: Facebook usage, size of online social network and number of good friends
Self-Reports	- Last reported energy/arousal and mood value and time - Energy/arousal and mood (average and variance) - Number of self reports

Table 5.2: User data and their parameters.

We also used the phone sensors and APIs to capture *contextual data* (see Table 5.3) Sensor data was collected during 5 seconds every 30 minutes. This was done to prevent battery drainage and in alignment with the operating system policies.

Sensor / API	Feature
Calendar	- Number of (free, not free) calendar records (before, during and after an intervention) - Time until the next meeting
GPS	- Number of records (at home, at work, null) - Time since GPR record at work - Signal quality (average, last record) - Location (distance to home, distance to work)\ - Distance traveled
Time	- Day of the week and Time - Lunch or Night time
Accelerometer	- X, Y, Z average, variance (jerk) – 30, 120 min Number of accelerometer records (30, 120 min)
Screen Lock	- Number of events - Time since last lock event

Table 5.3: Sensory features collected on the phone.

Output - Intervention Type Features

Five binary features were used to describe the type of intervention being selected. One feature was used as a signal to choose individual vs. social interventions and the other four features were used to select each of the four therapy groups (See Table 5.1).

Machine Learning Model

We trained ensembles of regression trees using the Random Forest algorithm [16]. After training, the model was capable of taking into account the user information to predict the expected reduction of stress after each intervention. We measured such reduction by calculating the delta between the subjective stress assessment (SSA) before and after the intervention. The Random Forest algorithm creates an ensemble of trees that are diversified by allowing each node to use only a subset of the available features. Since the goal of the model is to learn the differences between the interventions, the five intervention type features were given the higher probability to be enabled in every node.

Since we did not have data, we could not run batch comparisons between different algorithms. Therefore, we had to use the data from the first iteration as a seed. Random forest is a successful model, and it has the advantage that trees are easy to interpret. We wanted to use this practical meaning during the experiment to fine tune parameters. Our experience is that boosted trees are the best performing models for generic learning problems. However, we did not want to use boosted trees here since it is harder to incorporate the dynamic nature of this design. We could have chosen other models (neural networks, Gaussian processes, among others.) However, we wanted to use a reliable algorithm. The computational performance was not a big issue for this task.

Following the Upper Confidence Bounds (UCB) algorithms [7], we used optimistic predictions using the standard deviation computed from the deviation on the leaves of the trees that conform to the random forests of the average. Furthermore, we used this score to find the intervention that is expected to reduce stress as much as possible and/or tell us more about which micro-interventions to utilization in the future for such purposes. We retrained the Random Forest model on a daily basis mainly to avoid service failures. However, we did make incremental changes during the day to the scores on the tree leaves without changing the tree structure.

Featurization Strategy

We could not pre-select features since this is not a batch learning problem. Therefore, we had to use features that make sense. We tried to use features that will make sense such that we can interpret the results. Moreover, we gave different weights to the features randomly selected within the random forest algorithm. We required to have features describing the arm (the intervention) in a tree. Otherwise, this tree would not help in making the distinction between the different options. Accuracy is an important concern, but we could not test it up front since this is not a batch learning problem. Again, computational performance was not a big issue for this task.

Optimization parameters

We used the stress delta (stress level after the intervention minus before the intervention). This metric is a normalized value. People are different in their stress level and in the way they report their stress levels. People are better at comparing things than at assigning absolute values. Therefore, the delta is a more reliable signal.

5.3.3 Mobile App Interface

A mobile app was designed to interface with the user, deliver the interventions and gather input data for the ML algorithms. Additionally, an Experience Sampling Method (ESM) messaging interface was used to gather daily emotion self-reports.

App Design and Implementation

The design of the app was based on the following constraints: 1) support appropriate interactivity to drive engagement, 2) deliver content seamlessly, and 3) gather user and sensor data needed for the ML algorithms (see the following section). We chose a web- based format on Windows Phone v.8.1 with an embedded version of Internet Explorer v. 10 supporting HTML5. We used this system to track the metadata associated with the phone interaction and URL usage. We used a custom Azure cloud service to implement the data collection and user management modules.

User Flow

The app flow was designed using a dialogue-based schema. We presented the intervention prompts as a dialogue between the user and an agent (we chose an owl as a symbol of companionship and intelligence). We never used the word "intervention" in the dialogues, but rather the word "activity". The process followed five simple steps: 1) the user clicked on the app icon to request for an intervention; 2), the user was prompted to enter his/her current stress level (Figure 5.1a); 3) a micro-intervention (title, slogan and prompt) was presented (Figure 5.1b); 4) the interventions were all web-based and implemented in HTML5, so users simply needed to click on the play button to get to the URL page (or, depending on the condition, choose from a list of links suggested); 5) once the user experienced the intervention, they were asked to rate their stress level again.

Figure 5.1: Mobile app sample screens: a) subjective stress assessment; b) intervention title, icon, slogan and prompt.

During the first week of operation we observed a lot of unrealistic data (i.e., very high or very low stress reports with practically no time spent in interventions). We assumed this was due to people trying to take advantage of the incentives in place to encourage adoption. So we added a smaller footnote to the slider to elicit people's sense of duty (moral code) [87] in the hope this would encourage more practical use of the app (Figure 5.1a).

Experience Sampling Method (ESM)

ESM was used to track emotional variation during the day as an input to the ML algorithms. A two-dimensional Circumplex model of emotion [127] was used as the self-report rating interface. Users were prompted via a pop up message (Figure 5.2a) to use the Circumplex model (Figure 5.2b) and self report their emotional state approximately every 90 minutes (+/- 30 minutes) from 8am until 10pm. Users simply had to drag the circle to the quadrant that they felt they were in at that time (left to right for negative to positive valence, and bottom to top for low to high energy). A user who wanted to perform an intervention was not required to self-report their mood. However, if users wanted one intervention right after a self-report, they were prompted to do so (Figure 5.2c).

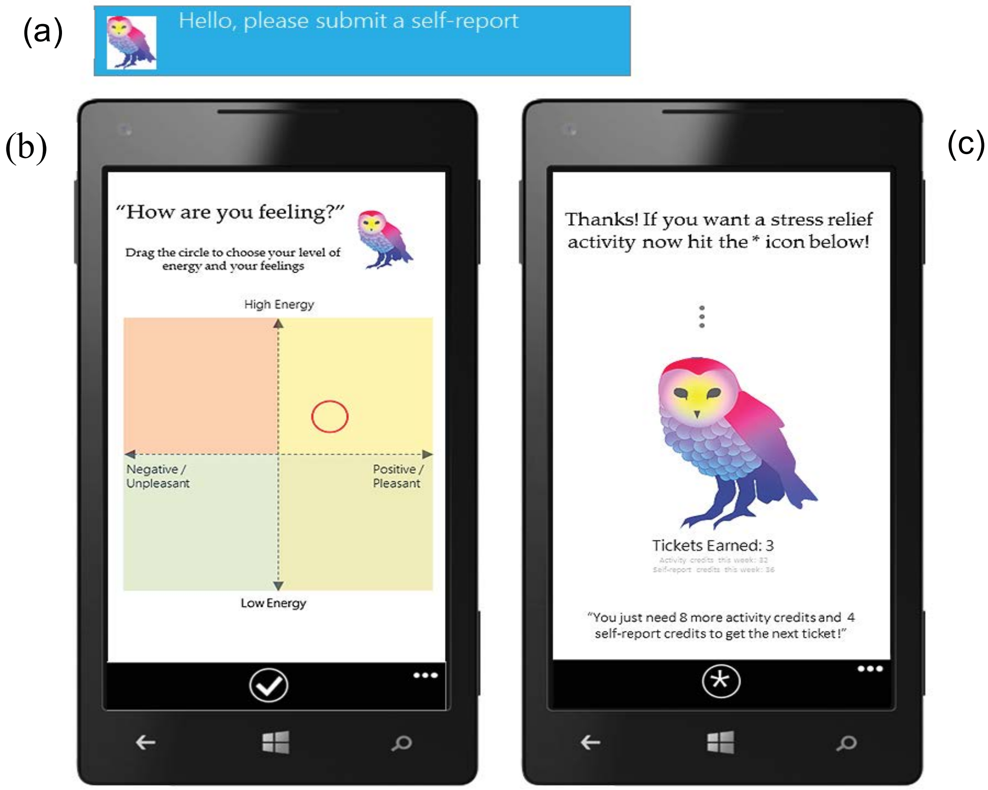


Figure 5.2: Experience Sampling Method (ESM): a) Self-report message to be clicked by user; b) Circumplex quadrant selection; c) Self-rating completion with option to launch micro-intervention if desired.

5.4 STUDY DESIGN

During the study we wanted to answer the three broad questions mentioned in the Introduction section. We chose four weeks as the length of our study for practical and monetary reasons. The study protocol was approved by the ethical and legal committee of the institution in which it was conducted. This section provides details about screening, experimental design and participation incentives.

5.4.1 Phones and Screening Procedures

Per design, our users had to own a Windows 8 phone. In addition, all of our users were screened to be between 18-60 years old, use social media and the web. We recorded information (but did not screen) about presence of any mental illness and also whether or not other family members had been diagnosed. After screening we ended up with 95 participants (25 women), with an average age of 30. As part of the initial recruiting process, we had participants fill out validated scales for: depression (Patient Health Questionnaire - PHQ-9) [73], coping with stress (Coping Strategies

Questionnaire - CSQ) [125], affective states (Positive Affect and Negative Affect Scale—PANAS short) [148] and gathered demographics info.

5.4.2 Experiment Design

The participants were divided into 4 groups for a 2 x 2 between subjects' experimental design: *ML v. Random Recommended Interventions* and *Self-selection from a Menu or Not* (users could take the recommended intervention offered or choose from a list). Table 5.4 shows the different conditions with the number of participants assigned to each category and the number of interventions performed at the end of the study.

	Random Choice	ML recommendation
Cannot self-select	22 users (23.1%) 1307 interventions (24%)	21 users (22.1%) 1176 interventions (22%)
Can self-select	26 users (27.4%) 1444 interventions (26%)	26 users (27.4%) 1550 interventions (28%)

Table 5.4: Distribution of participants and interventions for the different conditions of the study.

5.4.3 Gratuity and Incentive Policies

Participants opted into the study by accepting an email invitation after asserting that they would like to take part in the study. The email included a username and a password for downloading and installing our application, which was hidden in the Windows app store from the general public but available to our participants. Every week, for every 10 activities and 10 self-reports, each participant received a ticket to a weekly lottery of 3 x $100 gift cards; an additional ticket was awarded to participants who filled the weekly survey. On top of that, any participant that had at least 10 activities and 10 self-reports and had completed the survey on each of the 4 weeks was awarded a standard gratuity (~$300).

5.5 RESULTS

The study generated 26 days of data collection. First we present some descriptive statistics on interventions, stress deltas, drop out ratios, etc. Next, we present qualitative and quantitative results for the 20 users that used the app and filled out surveys for all four weeks, i.e. the group that completed the study in its entirety. Further qualitative analysis of the data from those users that did not complete the study is not included in this paper.

5.5.1 Descriptive Data

Recommendations and Selections

Figure 5.3 shows the distribution of the interventions recommended by the random recommender compared to the ML one.

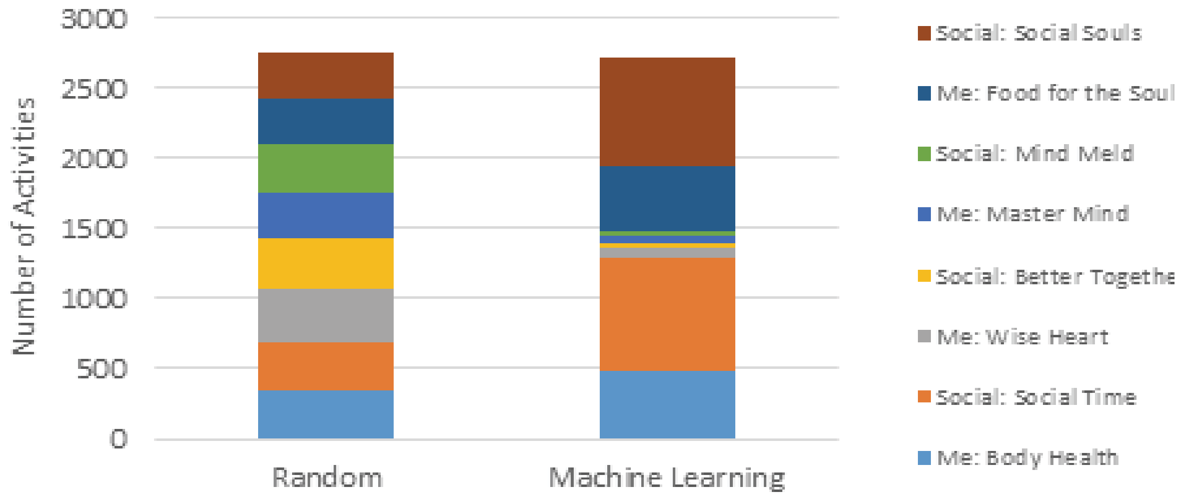


Figure 5.3: Machine Learning selected

Per design, the random recommender delivered uniform recommendations (320-380 times each), while the ML converged towards recommending mostly 4 types of interventions: Social Souls, Food for the Soul, Social Time and Body Health. Indeed, one should not expect the interventions to have equal benefits for all users, this is also demonstrated in the distribution of interventions for each of the participants in the ML groups (Figure 5.4).

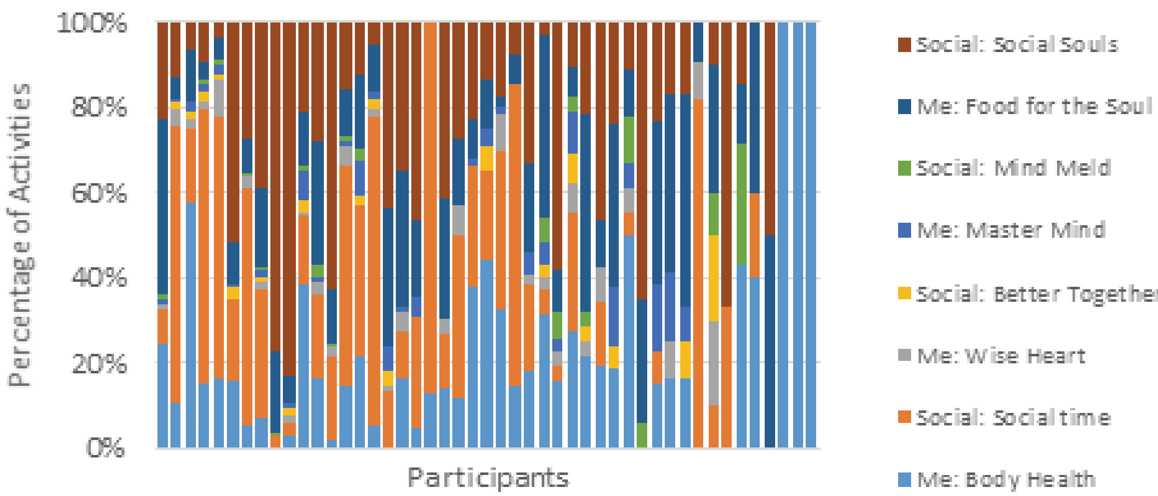


Figure 5.4: Distribution of interventions per participant.

The groups who could self-select used the recommended interventions the vast majority of the time, despite having the freedom not to. The ML group used the recommendations in 97% of the cases versus 98% of the time for the random group (See Figure 5.5). However, it seems as if during the last 10 days of the experiment, the participants in ML/self-select group used the selection option more often. This change towards the end of the study may be explained by seeking novelty effects when the ML became too “locked in” (i.e. stopped providing new types of interventions).

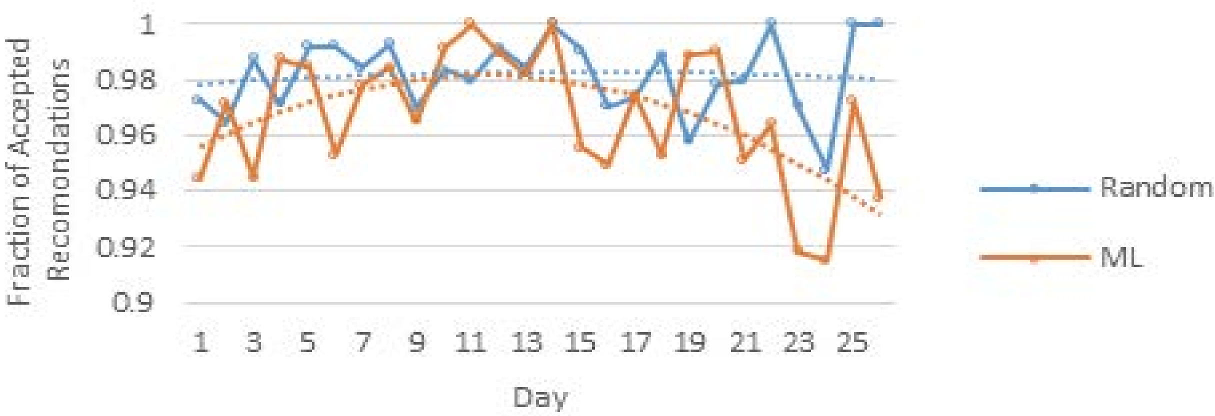


Figure 5.5: Fraction of interventions for which users selected the recommended intervention.

Stress Deltas

For each intervention completed, we have computed the delta between the stress reported before and the stress reported after the intervention. Figure 5.6 presents the average stress delta for the different groups on a daily basis. A paired t-test without the assumption of uniform variances was carried out for the post-pre stress deltas, $t(46)=2.06$, $p(\text{one-tailed})=0.02$. Users in the ML group reported significantly greater differences in stress reduction.

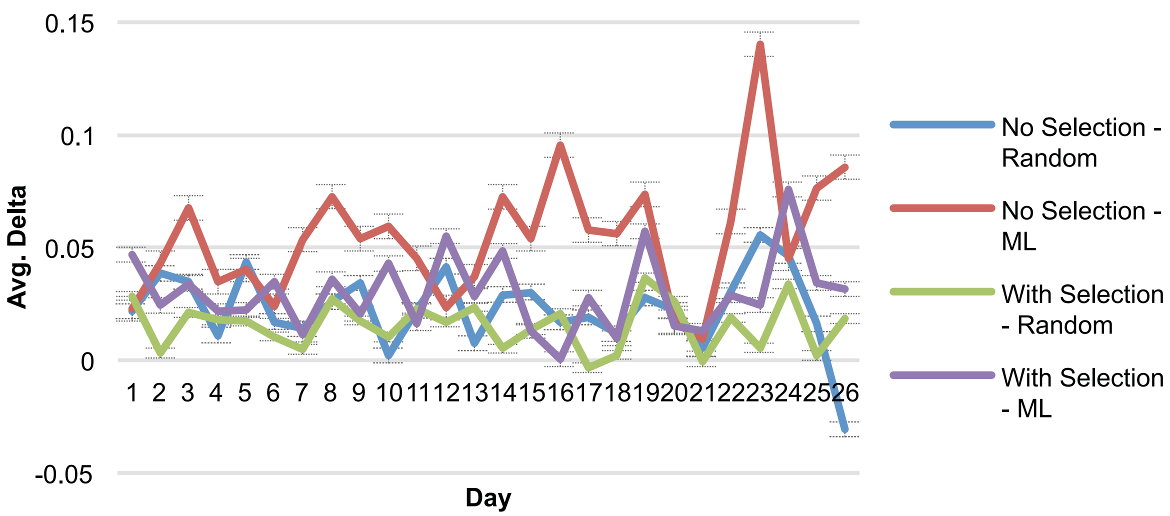


Figure 5.6: Stress deltas (stress before – stress after)

The group that received the machine learning interventions with no self-selection had the largest delta on most days. In particular, the average delta for this group (0.054) is greater than the other groups (0.016-0.021) and even greater than any single intervention (0.04). This shows that the use of the machine-learning algorithm increased the effectiveness of the interventions by doing better matching of the interventions to the participants and their context.

Drop-out

In terms of number of unique users per day, we saw a steady decline during the experiment; however, we did not notice large differences between the different groups (Figure 5.7).

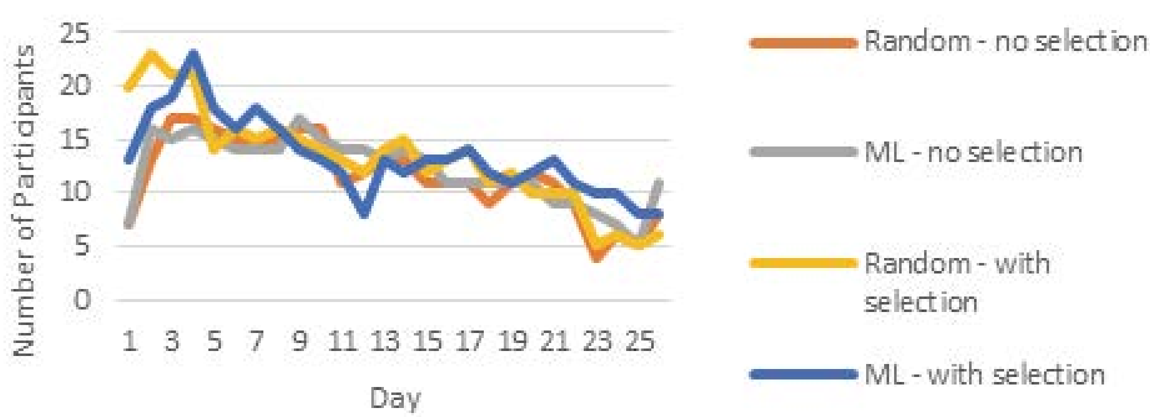
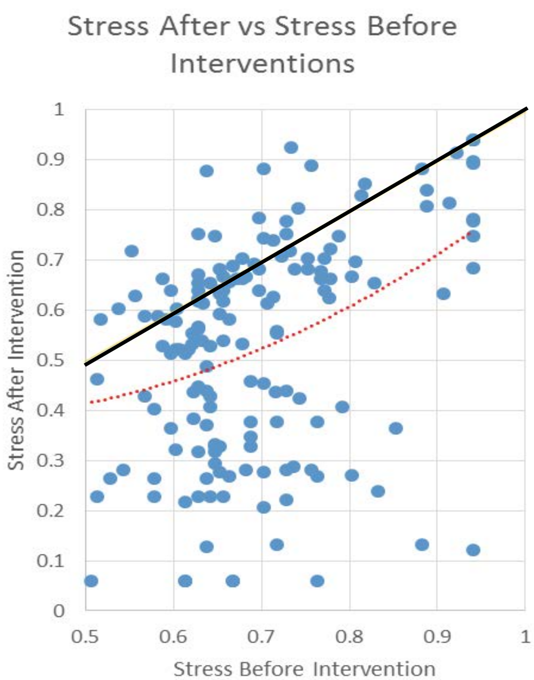


Figure 5.7: Daily users per day per experiment condition.

Correlational statistical analysis showed that significantly more users that were married, had children, had trouble at work or were sick were likely to drop out. There was no causal determination for this, obviously, but a future challenge will be to consider how people with so much stress and a busy life could be motivated to practice positive coping strategies. Maybe even going to an application on the phone was just too much for these users. At the end of the period we retained 21% of the population (N=20), with 10 users evenly distributed between the ML and random conditions and 12 users in the self-select condition versus 8 users in the non-self-select condition.

"Ideal" Users

During the study introduction we prompted users to use the app whenever they felt stressed. So, we were intrigued by a couple of "abnormal" behaviors: a) extremely short intervention usage time (< 3 sec) with very high (~1) or very low self-rating scores (~0); and b) reporting low pre-intervention stress levels (< 0.5), i.e., maybe not being stressed in the first place. We believe that part of this behavior could be explained by users trying to take advantage of the incentives. We observed "ideal" users, i.e., those with stress>0.5 and using interventions for more than 60 seconds. Figure 5.8a shows a higher stress reduction with a higher stress precursor.



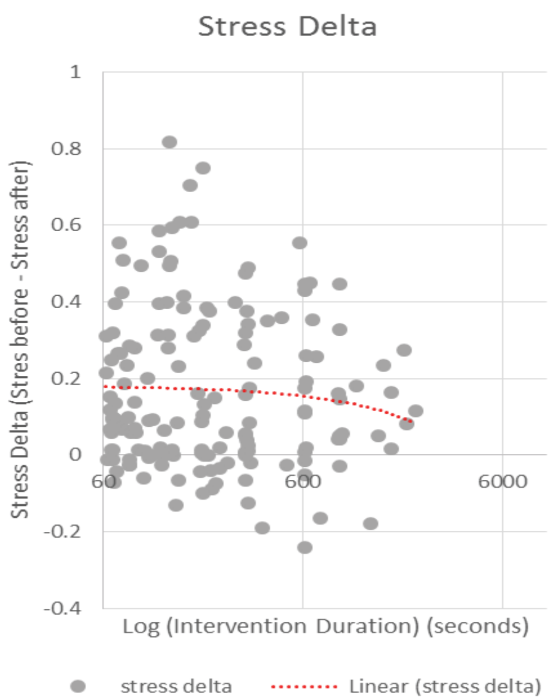


Figure 5.8: "Ideal" users (stress self-rating before intervention > 0.5 and intervention duration > 60 sec) a) Stress after versus stress before interventions; b) stress delta versus intervention duration.

The black line represents the “no effect” line, i.e., stress delta = 0. The red line shows the average stress delta. Users that were not stressed reported lower gains in terms of stress relief (avg. delta = 0.037) than stressed people (avg. delta =0.096). We ran a 2 x 2 (ML or Random vs. stress <0.5 or stress >0.5) RM- ANOVA. There was a significant effect of having stress >0.5 before an intervention, $F(1,18)=21.6$, $p<0.001$. No other effects or interactions were significant. In other words, having a stressful precursor resulted in a three times larger reduction of stress after performing a micro-intervention. Additionally, Figure 5.8b shows that interventions with a usage time larger than ~200 sec offered diminishing results. This is an interesting marker for the suggested optimal length of the interventions of the type we chose to use.

5.5.2 Qualitative Results

We asked a series of questions to gather information about user’s interaction with the interventions and their learning process.

Subjective Data

We obtained ratings of what users considered to be the intervention they liked and disliked the most, and the interventions they thought were effective or ineffective. Figure 5.9 shows the comparison of each aggregated number of counts for all the weeks.

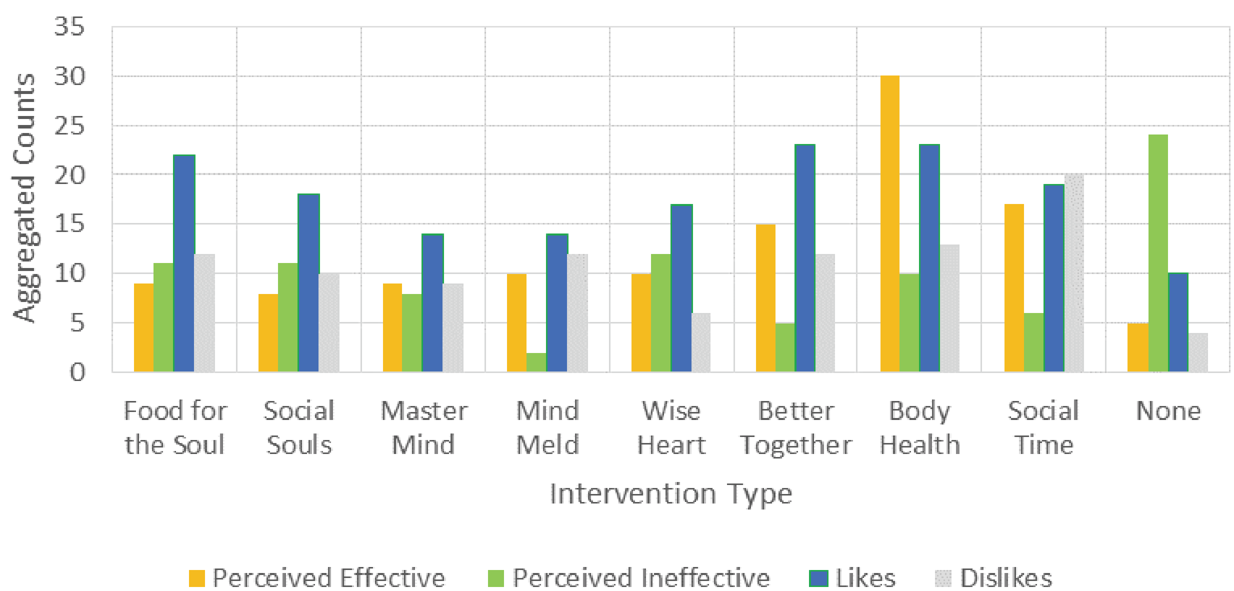


Figure 5.9: Subjective ratings (likeability and efficacy)

Clearly, Body Health (somatic-individual) received the highest scores for effectiveness and likeability, closely followed by Food for the Soul (positive psychology-individual), Social Time (somatic-social), and Better Together (meta-cognitive- social). The positive psychology groups, Food for the Soul, and Social Souls, although liked, received low grades on perceived efficacy. It is also interesting to note that the most rated interventions align with the ML recommendations.

Learning and Awareness

Table 5.5 shows the answers to the question *"What have you learned from this study?"* 70.3% of the users reported a higher stress self-awareness; however, it is interesting to observe that 34% reported stress awareness as stressful. 65.6% reported having learned simple ways to control stress. A paired t-test without the assumption of uniform variances was carried out for the post-pre stress deltas, $t(46)=2.06$, $p(\text{one-tailed})=0.02$. Users in the ML group reported significantly greater differences in stress reduction.

"What have you learned from this study?" (Multiple choice question)	%
To be more aware of my stress levels	70.3%
Simple ways to control my stress	65.6%
That being more aware of my stress level is stressful	34.4%
Nothing	7.8%
Other	4.7%

Table 5.5: Reported learning.

Most users reported higher levels of stress awareness due to the use of the app. Comments like: *"Although I did not do a good job of using the app this week, by using it in weeks past I am still aware of when I become stressed and try to deal with it"* or *"Doing the study helped me spotlight it"* showcase the way people extrapolated the benefits of the study beyond the use of the app.

A number of participants also reported having learned that simple methods can help manage stress. Comments like: *"I breathe and take time for myself to clear my mind"* or *"(I) take time to take care of my body and soul"* showcase the way some people found inspiration in micro-interventions to do something about stress.

5.5.3 Quantitative Results

There were only 20 participants left at the end of the four-week study period. These were participants who followed all of our instructions to use the app multiple times per week, including the self-ratings and interventions, and completed the end of the week surveys every week. Some participants told us that they only needed 1 or 2 weeks to learn the intervention skills.

Nonetheless, to objectively look at the benefits of using the app for the whole 4 weeks, we needed to use this set of participants who made it the full way. Deep analysis of why and which participants dropped out after the first few weeks remains future work.

Depression – PHQ9

The PHQ-9 response data was analyzed for the 20 participants who used the app all 4 weeks. A 2 (ML or not) x 2 (Selection or Not) RM-ANOVA was carried out on the average score values for the initial survey week (pre-application baseline) and the 4 weeks after using the app, for 5 replications. A significant effect of week, $F(4,76)=2.9$, $p=0.026$, was found, and ML was borderline significant, but no effect was observed for the Selection variable. When the data was collapsed across the Selection variable, there was a borderline significant effect interaction for week x ML ($p=0.06$). This means that, regardless of ML condition, participants showed statistically significantly less depression level (DL) while they used our tool. In addition, the ML group added to this improvement more than Random selection (borderline). In clinical terms, ML condition users ended week 4 with no signs of depression (DL < 5), while the Random condition ones showed mild depression symptoms (5 < DL < 10) [73] (Figure 5.10).

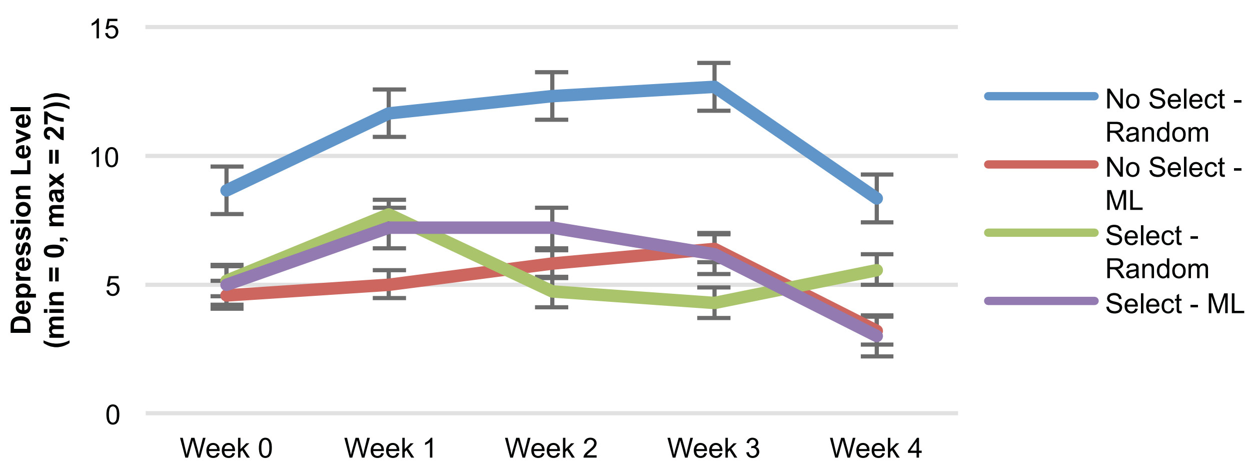


Figure 5.10: Depression (PHQ-9) quantitative results.

Coping – CSQ

A 2 (Random v. ML) x 2 (No selection v. selection from a menu) x 4 (week) RM ANOVA was performed on the differences between constructive and destructive coping behaviors to see if our participants were *learning and incorporating* new coping strategies via our interaction tool. A significant 3 way interaction was observed, $F(1,16)=4.4$, $p=0.003$. No other significant effects emerged from the analysis. While a 3-way interaction can be difficult to understand, as observed in Figure 5.11, the group with ML without selection reported significantly more constructive coping behaviors over time. This is an encouraging finding as it indicates that those users were willing to trust the personalized intervention offered by the ML algorithm by using it and indicating greater stress relief.

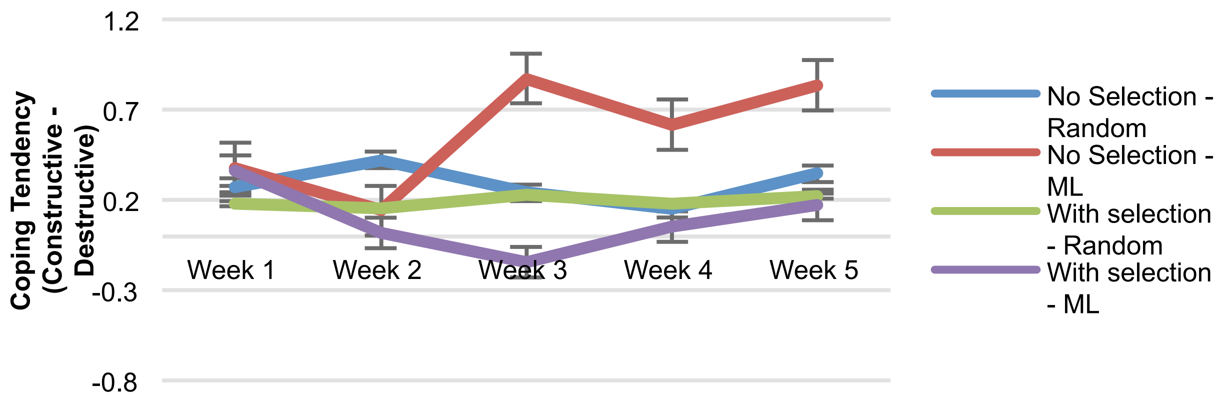


Figure 5.11: Coping Strategies Questionnaire (CSQ)

Overall, results support the hypotheses (see Intro); a *recommender* system was able to deliver a suite of *web apps* that helped people reduce stress locally and *learn to cope* with stress over time.

5.6 IMPLICATIONS FOR DESIGN

Several design implications were identified during the study and can be presented in two groups: intervention design and intervention recommendation.

5.6.1 Intervention Design

Simplicity & Small Size

Our simple and short intervention format (short prompt + familiar URL) allowed people to use the app in different contexts. Furthermore, this format can scale up to other interventions that may work better for different people and/or situations. More complex formats have a higher risk of reduced adoption and adherence. Further work may focus on the optimal intervention duration depending on the person and the context.

Incentives

A limitation of a usage-based incentive system is that users end up using the system without needing it. One way to overcome this is to make the app public so that no incentives are needed for its use. The limitation of this model is that an advertisement campaign should be properly crafted to drive initial awareness and adoption.

Stress Awareness

As reported, stress awareness was a driver for people's use of the application. However, it was itself a source of stress to 1/3 of the users. This is a limitation of ESM as a source of data. Tailoring self-reporting frequency, or even eliminating its use, should be considered when developing stress management systems.

5.6.2 Intervention Recommendation

Ensuring Novelty

Our experiment suggests that using ML helps in matching interventions to the user's context and hence, improves the overall outcomes for users, in terms of stress level, depression and coping (for the non-selection users). There are several ways in which the ML algorithm can be improved. For example, the algorithm reduced the diversity of the interventions sent to the participants. This might have resulted in boredom, and in the long run, might lead to a high attrition rate. This may be improved by increasing the number of interventions in each group and adding diversity as an objective to the ML algorithm. Another approach would be to use periodic surveys to update participants' models, which can lead to changes in the types of interventions presented.

Exploration vs. Exploitation

The ML algorithm addressed the problem of exploration (searching for new options) vs. exploitation (refining existing choices); however, the group that was given random interventions did most of the exploration. In a sense, the design of the experiment dictated that, for at least 50% of the time, the model was exploring. This was needed to validate the use of the ML matching algorithm as an intervention selection tool. However, now that we have validated this conjecture, in future studies, one may not wish to use 50% of the interventions for exploration.

Targeting "Ideal Users"

As described, targeting "ideal" users, i.e., people aware of their need for a stress management recommender and who are able and willing to use it should increase the local effect of interventions. ML algorithms could adapt weights for these users' inputs. Targeting populations that need stress management could teach us more about the efficacy of the recommender system and the interventions themselves. However, the challenge remains to create systems that help prevent stress in the general population.

5.6.3 Future Research

As mentioned in our introduction, the challenge to deliver effective interventions in real life was framed as: how can we design the "right" intervention(s) to be delivered at the "right" time(s)? Many interesting questions still remain in terms of "what" interventions should be delivered. In a new iteration of this system, we plan to explore the authoring problem. We want to explore crowd and self-authoring as direct sources of new interventions and social media data mining as an indirect source. We will further explore the types and duration of the interventions as a factor of adoption, as well as a new variation of the app based on complementary qualitative analysis of elements such as the mascot, the interaction with the experience sampling method, the flow, among others. With regards to "when" is the right moment to intervene, we plan to do experiments where we use psycho- physiological sensors to trigger the interventions. We want to study not only if the sensors can determine the best time to intervene, but also if they drive awareness and motivation in users.

5.7 CONCLUSION

In this chapter we have shown the potential for popular web apps to provide an "unlimited" source of not only inspiration, but also actual stress management interventions. We showed that ML algorithms could be used to improve engagement and local efficacy by matching the right intervention to the context of the user. Finally, we observed a tendency from users to adopt constructive coping strategies, not only by using the interventions suggested, but also by understanding that simple activities can actually help them to manage their stress. We find these results encouraging with regards to continuing research to enable "popular" therapies, mechanisms to assist large populations to cope with daily stress and drive sustained behavior change.

Chapter 6
Multi-modal Interventions

Little is known about the affective expressivity of multisensory stimuli in wearable devices. While the theory of emotion has referenced single stimulus and multisensory experiments, it does not go further to explain the potential effects of sensorial stimuli when utilized in combination. In this chapter, we present an analysis of the combinations of two sensory modalities: *haptic* (more specifically, vibrotactile) stimuli and *auditory* stimuli. We present the design of a wrist-worn wearable prototype and empirical data from a controlled experiment (N=40) and analyze emotional responses from a dimensional (arousal + valence) perspective. Differences are exposed between “matching” the emotions expressed through each modality, versus "mixing" auditory and haptic stimuli each expressing different emotions. We compare the effects of each condition to determine, for example, if the matching of two negative stimuli emotions will render a higher negative effect than the mixing of two mismatching emotions. The main research question that we study is: When haptic and auditory stimuli are combined, is there an interaction effect between the emotional type and the modality of the stimuli? We present quantitative and qualitative data to support our hypotheses, and complement it with a usability study to investigate the potential uses of the different modes. We conclude by discussing the implications for the design of affective interactions for wearable devices.

6.1 INTRODUCTION

As wearable devices (a.k.a. "wearables") become increasingly equipped with multiple sensors and actuators, it is expected that perceptual experiences will be influenced not only by how each of these sensors and actuators are used individually, but also by their various combinations. Individual audio signals have been studied extensively as carriers of emotional content [15, 155]. Hertenstein et al. [56] have also shed light on the emotional expressivity of haptic signals. However, the affective response of these sensory stimuli is understudied.

This chapter examines such multisensory combinations. More specifically, we study the perception of emotions when triggered by sounds and/or vibrotactile stimuli. We choose a wrist-worn device mainly due to its large adoption as a wearable, despite the complexity of generating haptic stimulus due to its limited contact with the human body.

We evaluate emotion from a dimensional perspective, based on the continuous measurement of their emotion components: arousal and valence [56, 127]. In general, we expect that the combination of multiple stimuli should generate different

responses depending on the intensity and the value of their arousal and valence components. Our main research question is: *When haptic and auditory stimuli are combined, is their combination linear in nature?*

A preliminary qualitative assessment helped illuminate the nature of emotionally non-matching haptic and audio stimuli combination. The combination of non-matching stimuli (i.e. pertaining to two different quadrants) generated responses of surprise and disgust, which indicate that results will most likely fall in the high arousal, low valence quadrant. With this information, we formulated the following hypothesis:

> *H. When combining sound and vibrotactile stimuli, the following interaction effects will be observed:*
>
> > *H.1 auditory stimuli will be dominant in both axes (valence and arousal).*
> >
> > *H.2 haptic stimuli will modify (reduce or enhance) the effect of sound.*

To test this hypothesis, we built a wearable prototype made of two eccentric rotating mass (ERM) vibration motors to induce two haptic phenomena known as "apparent tactile motion" and "phantom tactile sensation" [63], in order to generate some of the affective gestures described by Hertenstein et al. [56]. We performed a controlled lab experiment with 40 users (52% female and 48% male) balanced across conditions. We controlled the order in which the stimuli were presented to the user. Our results show that there is no interaction between stimuli modes and their emotional types (arousal / valence). Single-mode stimuli showed a higher valence for sound (vs. vibrotactile). Multimodal interactions did not reflect a clear dominance of sound, which was modulated by haptic stimuli. However, the lack of clear emotional expressivity of the single modes stimuli affected the combined modes. Furthermore, qualitative data supported these findings.

6.2 BACKGROUND

6.2.1 Circumplex Model of Affect

The Circumplex Model of Affect (CMA) by Russell [127] essentially represents emotions as a combination of two dimensions: arousal (ranging from sleep to high arousal) and valence (ranging from displeasing to pleasing). As depicted in Figure 6.1, emotions are organized in a two-dimensional space. The horizontal axis represents the feelings/valence (pleasure- displeasure), while the vertical axis represents the degree of arousal (excited-sleep). When looking at the CMA, it becomes apparent that each of the four different quadrants contains similar emotions.

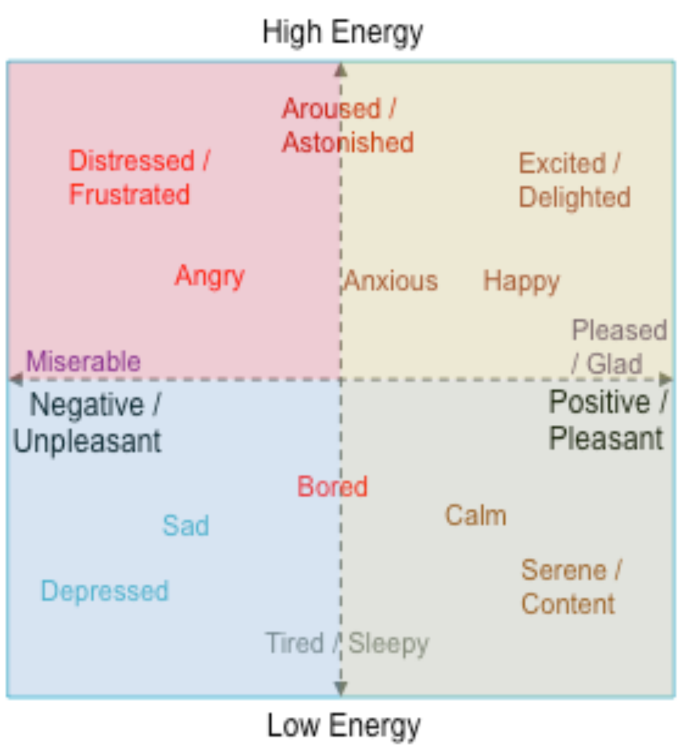


Figure 6.1: Circumplex Model of Affect by Russell [16].

The upper right quadrant is themed around emotions of excitement or happiness, the upper left represents distress or anger, the lower left is representative of depression or sadness, and the lower right revolves around feelings of contentment or relaxation.

6.2.2 Auditory Affective Expressivity

Affective expressivity of sounds is well studied and documented in the International Affective Digitized Sounds (IADS-2) library developed by Bradley and Lang [15]. The IADS-2 is a set of standardized, emotionally-evocative, internationally accessible sound stimuli that consists of a total of 167 sounds, each of which is characterized by its mean values for arousal and valence.

6.2.3 Haptic Affective Expressivity

A comprehensive study of affective expressivity of human touch, performed by Hertenstein et al. [56] evaluates twelve emotions divided into three groups: (a) 6 emotions based on Ekman's traditional studies of emotion [36] (anger, fear, happiness, sadness, disgust, and surprise), (b) 3 prosocial emotions (love, gratitude, and sympathy) [134], expected to have higher expressivity via touch, and (c) three self-focused emotions (embarrassment, pride, and envy) [154], which play as counterparts to the prosocial ones. In addition, by observing how participants expressed emotion through touch, Hertenstein et al. [56] could determine the types of touch people used to communicate specific emotions. Table 6.1 shows a subset of the types of touch selected for this chapter.

Emotions	**Types of touch – in order of relevance**
Anger	Hitting, Squeezing, Trembling.
Fear	Trembling, Squeezing, Shaking.
Happiness	Swinging, Shaking, Lifting.
Sadness	Stroking, Squeezing, Lifting.

Table 6.1: Top four haptic patterns used for the expression of emotions[9].

6.2.4 Multisensory Affective Response

Haptics and visual stimuli interaction has been recently studied showing a dominance of the visual stimulus, despite the different modulations of the haptic signals performed by the authors [1]. No additional work in other types of multimodal affective responses have been reported to date.

6.3 PRIOR WORK

6.3.4 Audio and Haptic Research

Haptic and audio interaction workshops [119] conducted over the course of the past decade have studied the use of audio and haptic interaction in fields as diverse as music, interfaces, and communications [18, 139, 145], with a focus on the design and evaluation of novel multimodal interactions. However, the use of haptic technology to support affective computing is barely studied. Within the literature, we find that haptic stimulus could help reduce the amount of sensorial or cognitive load when implementing calming technologies [112, 117]. Another line of study is the design of frameworks to examine the use of haptics primitives such as distance, surface contact, time exposure, movement, and oscillations, among others [28, 46].

6.3.5 Vibrotactile Interfaces

Israr and Poupyrev [63] explored the use of two vibration sources to create the illusion of apparent tactile motion, which could be used to mimic the touch of a human stroke, or the phantom tactile sensation, which could be used to mimic a poke (see Figure 6.2). Recently, Huisman et al. [61] evaluated interactions based on a forearm sleeve used to express certain movements such as a poke, a hit, pressing, squeezing, rubbing, and stroking. Richter et al. [124] explored scenarios where the phantom tactile sensation can be recreated using different actuators with direct applications in ubiquitous computing.

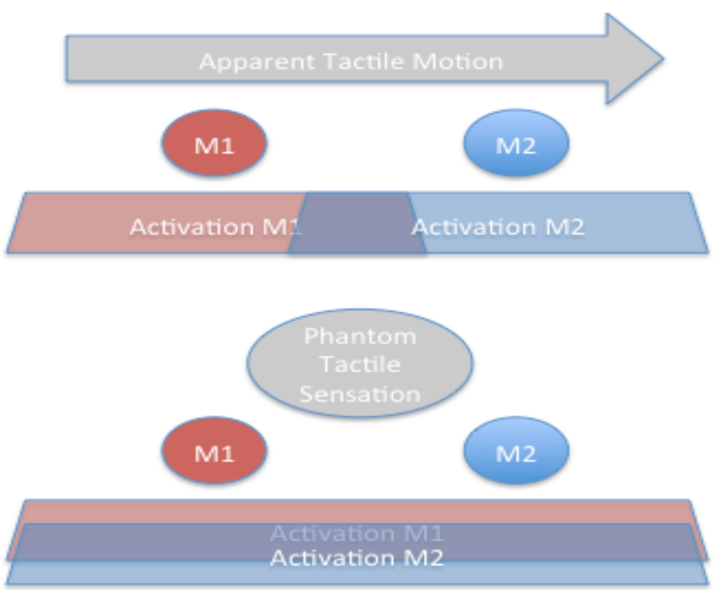


Figure 6.2: Basic haptic effects with two vibrating motors as described in [12].

6.4 EXPERIMENT DESIGN

6.4.4 Vibrotactile Wearable Hardware

We built a wristband made of two Velcro strips. In between the strips, we fixed two small 5V DC ERM vibration motors with a variable speed ranging from 0 to 9000 RPM separated 1.1 inches apart and oriented to be parallel to the forearm bones. Figure 6.3a shows the bracelet connected to a pulse-width modulation (PWM) pin of the microcontroller. Figure 6.3b shows the placement of the motors on a human forearm.

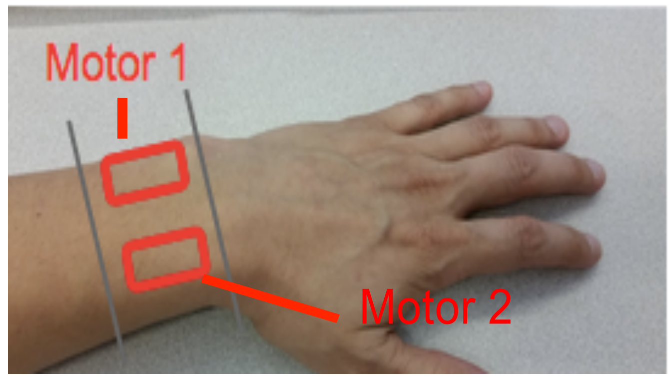


Figure 6.3: a) Prototype wristband made of Velcro straps covering the vibrating motors and attached to Arduino Uno. b) Motor placement (1.1 inches apart).

We recreated an “apparent tactile motion” effect [63] by using an overlap time of 200ms between the end of the stimulus in one motor and the beginning of the next one. We recreated a “phantom tactile sensation” effect [63] by activating both of the motors in parallel.

6.4.5 Stimuli Generation

Table 6.2 shows a representation of the CMA with the selected vibrotactile and sound stimuli for each quadrant. We gave them short labels (happy, relaxed, sadness, anger) for easier identification purposes. Sounds were played on a speaker under the chair’s armrest at a uniform intensity, which was calibrated to the user’s comfort, while vibrotactile intensity depended on the type of interaction. The stimuli had a duration of 6.5 seconds, which seemed to be a good time to elicit enough emotional information based on our previous pilot observations.

We selected our touch patterns based on the touch primitives described by Hertenstein et al. [56]. We chose a "hitting" touch to represent the high arousal + low valence (anger) quadrant, a "swinging" touch for the high arousal + high valence (happiness) quadrant, and a "stroking" touch for the low arousal + low valence (sadness) quadrant. Hertenstein et al. [56] did not research emotions relating to the low arousal + high valence quadrant (relaxation), so we chose a low intensity long touch based on the idea of a "relaxing" massage.

Haptic: Hitting → Hit High Intensity (5V) Short Burst (100ms both motors) “Phantom Tactile Sensation” **Sound → Alarm (IADS)** Valence = 4.3, Arousal = 6.99 Short label: “Anger”	**Haptic: Swinging → Strokes** Moderate Intensity (4V) Left-Right and Right-Left fast strokes (190ms, 50ms overlap) “Apparent Tactile Motion” **Sound: People Laughing (IADS)** Valence = 7.78, Arousal = 5.942 Short label: “Happy”
Haptic: Stroking → Stroke Moderate Intensity (4V) Left-Right slow stroke (380ms, 100ms overlap) “Apparent Tactile Motion” **Sound: Violin (own selection)** Valence = N/A, Arousal = N/A Short label: “Sadness”	**Haptic: Massage* → Press** Low intensity (2.5V) Long vibration (1500ms) “Phantom Tactile Sensation **Sound: Harp (IADS)** Valence = 7.44, Arousal = 3.36 Short label: “Relaxed”

Table 6.2: Circumplex Model of Affect with haptic and sound mappings per quadrant.

We translated our selected touch patterns into vibrotactile stimuli based on the implementations described by Huisman et al. [61]. Touch patterns were created based on different combinations of the duration and the intensity of vibration. The “hitting” touch was represented by a poke using the "phantom tactile sensation" effect with a short duration and high intensity vibration. The "swinging" touch utilized the "apparent tactile motion" effect, by using quick left to right and right to left strokes of moderate intensities. The "stroking" touch also used the "apparent tactile motion" effect with a unidirectional stroke of moderate intensity. Lastly, the "massage" touch used a long duration and low intensity activation of both motors. We selected the sounds from the IADS-2 database based on the average arousal and valence values.

Unfortunately, IADS-2 does not have a good low valence + low arousal sound. So, we queried an open source sound database [44] for a “sadness sound” and settled with the sound “Sad Violin.”

6.4.6 Conditions

We conducted an experiment focused at evaluating the interaction between single and multimode stimuli as well as the differences between haptic and auditory-only stimuli. As it can be seen in Table 6.3, in essence we designed a 2x2 counterbalanced experiment to reduce ordering effects.

	Single Mode (first)	**Combined Mode (first)**
Haptic (first)	**A)** 2x 4-Haptic stimuli 2x 4-Sound stimuli 2x 16-combined stimuli	**B)** 2x 16-combined stimuli 2x 4-Haptic stimuli 2x 4-Sound stimuli
Auditory (first)	**C)** 2x 4-Sound stimuli 2x 4-Haptic stimuli 2x 16-combined stimuli	**D)** 2x 16-combined stimuli 2x 4-Sound stimuli 2x 4-Haptic stimuli

Table 6.3: Experiment conditions

Participants were asked to assess arousal (low to high energy) and valence (unpleasant to pleasant feelings) at the beginning of the experiment and after each stimulus using Likert scales from 1 to 9. The single-mode blocks consisted of 4 different auditory or 4 different haptic stimuli, all randomized. The multimodal blocks consisted of 16 randomized stimuli made of the combination of the 4 vibrotactile and the 4 sound stimuli. After each block there were short 30 second breaks. At the end of the experiment, they were asked to provide open text qualitative feedback about their experience and complete a usability questionnaire.

We gathered two (repeated) runs for each condition in order to reduce novelty effects. Weierich et al. [150] describe the effects of novelty in the amygdala as being very similar to those triggered by stimuli with high (vs. low) arousal and negative (vs. positive) valence responses. A simple average of the values of the two (repeated) runs can reduce novelty effects by improving the estimated value of each metric.

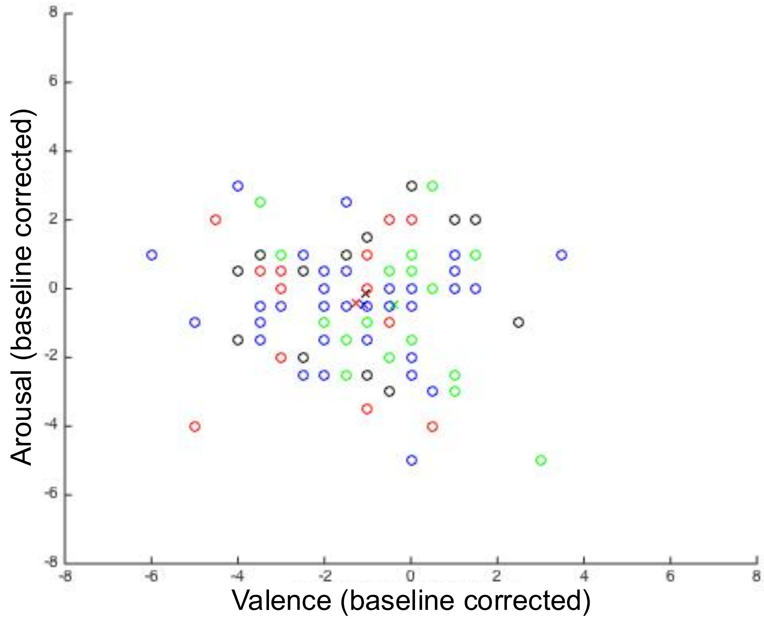
6.5 RESULTS

6.5.1 Quantitative Analysis

Our quantitative analysis follows a parsimonious procedure. It starts with a three-way hierarchical (nested) ANOVA that compares stimulus *mode* (haptic vs. auditory), emotional *type* (Low/High valence vs. Low/High arousal) and emotional *match*, a nested variable that differentiates between stimulus types (same valence and arousal levels) versus non-matching stimulus types (different valence and arousal levels) for the combined conditions. We created two models, one for arousal and one for valence.

Pre-processing and Data Visualization: To compare our Likert scale measurements, we normalized the intercepts across subjects by subtracting the baseline value measured at the beginning of the study. The arousal and valence scales that ranged from 1 to 9 were converted to scales that ranged from - 8 to 8 with a common zero value. Figure 6.4 shows the mapping onto the CMA of the haptic (a) and sound (b) stimuli values respectively.

a)



b)

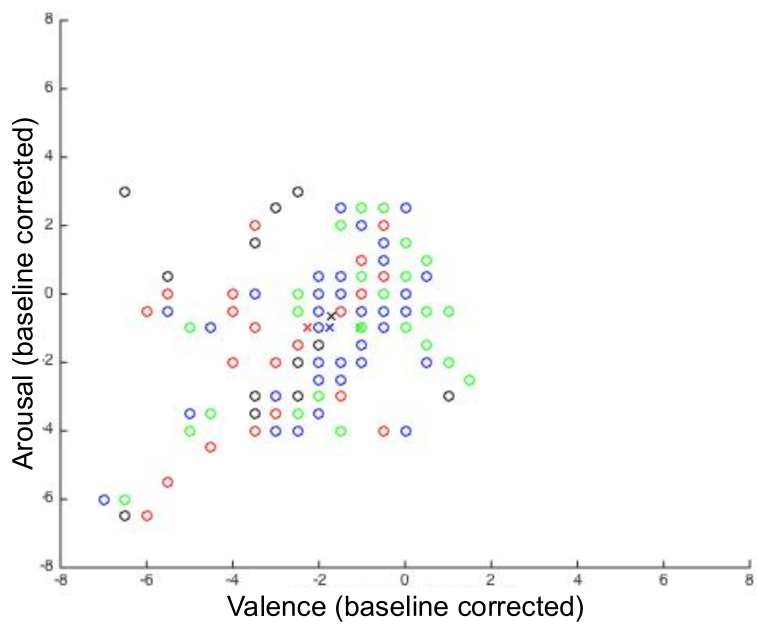


Figure 6.4: Stimuli affect map: a) Haptic, b) Sound. (o) represents average points for all 40 users; (x) represents the centroids; colors represent the different stimuli: black = happy, red = anger, blue = sadness, green = relaxed.

As it can be observed, there is little difference between the centroids and its distributions, and most of the values were below zero for both arousal and valence. Having most of the values below zero indicates that in general, people did not find the experience exciting (low energy/arousal) and that it generated some discomfort (unpleasant feelings).

Arousal Data Analysis: Results from a three-way hierarchical (nested) within-subjects ANOVA for mode, type and match(type). No interaction effect was found. A two-way analysis ANOVA without the match (nested) effect rendered no interaction effects $F(3,897) = 0.98$, $p=0.4046$. Main effects were observed for Type $F(19,897) = 2.03$, $p=0.0072$. Table 6.4 shows the results. A post-hoc Tukey multiple comparisons test revealed that a Relaxed-Relaxed (M=0, SD=0.1663) were marginally different to a Sad-Sad (M=-.0825, SD=0.1663) p=0.0562 and to a Happy-Happy (M=-0.825, SD=0.1663).

Source	**Sum Sq.**	**d. f.**	**Mean Sq.**	**F**	**Prob>F**
Type	45.44	19	2.5243	2.03	0.0072
Mode	0.05	1	0.05	0.03	0.8575
Type*Mode	2.88	3	0.9583	0.98	0.4046
Error	0	0	0		
Total	2834.16	959			

Table 6.4: Two-way ANOVA for Arousal values.

A larger difference between matching pairs seems to indicate a stronger reaction to coherence. It is not easy to explain why people report higher arousal to matching relaxation stimulus. H.1 cannot be validated. However, there seems to be some indication H.2 could be accepted with more data.

Valence Data Analysis: Results from a three-way hierarchical (nested) within-subjects ANOVA for mode, type and match(type) showed no interaction. A two-way within subjects ANOVA, shown in Table 6.5 reflects an interaction effect between Type and Mode $F(3,897) = 5.17$, $p=0.0022$, as well as main effects for Type, $F(19,897) = 4.7$, $p=0$ and for Mode, $F(1,897) = 6.93$, $p=0.0121$.

Source	**Sum Sq.**	**d. f.**	**Mean Sq.**	**F**	**Prob>F**
Type	166.42	19	8.7587	4.7	0
Mode	23.11	1	23.1125	6.93	0.0121
Type*Mode	16.36	3	5.4542	5.17	0.0022
Error	0	0	0		
Total	3157.42	959			

Table 6.5. Three-way ANOVA for Valence values.

Multi-comparisons for Mode revealed a higher valence (M=-0.78, SE=0.1417) than haptic (M=- 1.32, SE=0.1417) or combined (M=-1.34, SE=0.1417). This hints a potential dominance of sound over other modes.

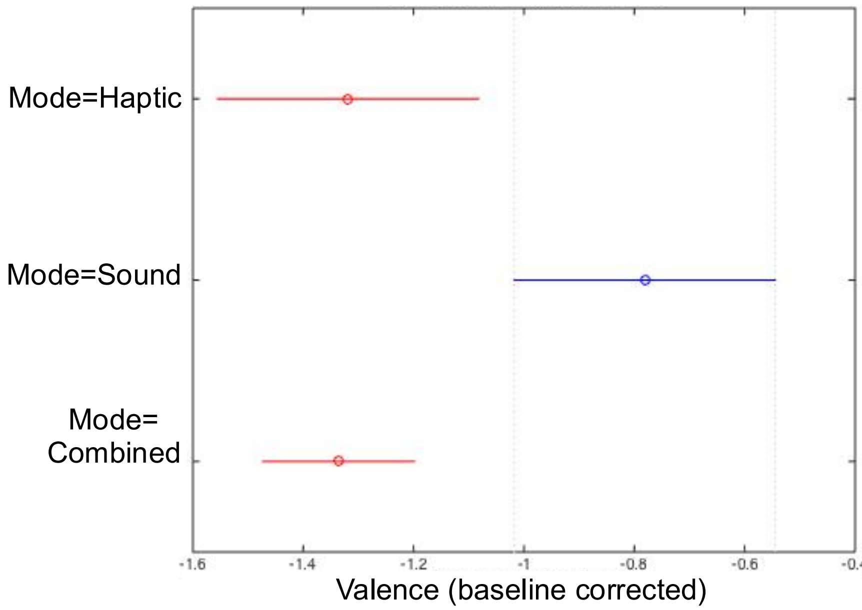


Figure 6.5: Multiple comparisons (with Bonferroni's correction). Sound mode's valence is significantly different than Haptic mode or Combined (Haptic + Sound) mode.

Figure 6 shows a few multi-comparisons for Type of the different Haptic+Sound pairs. Several interesting differences are revealed. First of all, all combinations with an Angry Sound had lower valence than the matching Relaxed(haptic)-Relaxed(sound) combination (M=-0.5, SD=0.214) and the single Relaxed mode (M=-0.6875, SD=0.1213) (Figure 6.6a). All combinations of a Relaxed sound and all the single modes except Angry had higher valence than the Happy(haptic)-Angry(Sound) combination (M=-2.05, SD=0.214) (Figure 6.6b). The Angry(Haptic)-Happy(Sound) (M=-1.8, SD=0.214) had lower valence than the Relaxed single mode (M=-0.6875, SD=0.1513) and any other combination of the Relaxed sound, except the Sad-Relaxed (M=-0.7, SD=0.214) one. These results seem to indicate that the Relaxed sound and its matching haptic did have more soothing effects than other combinations. Furthermore, a highly opposing combination such as Happy (sound) and Angry (haptic) seem to generate much more negative feelings than other effects.

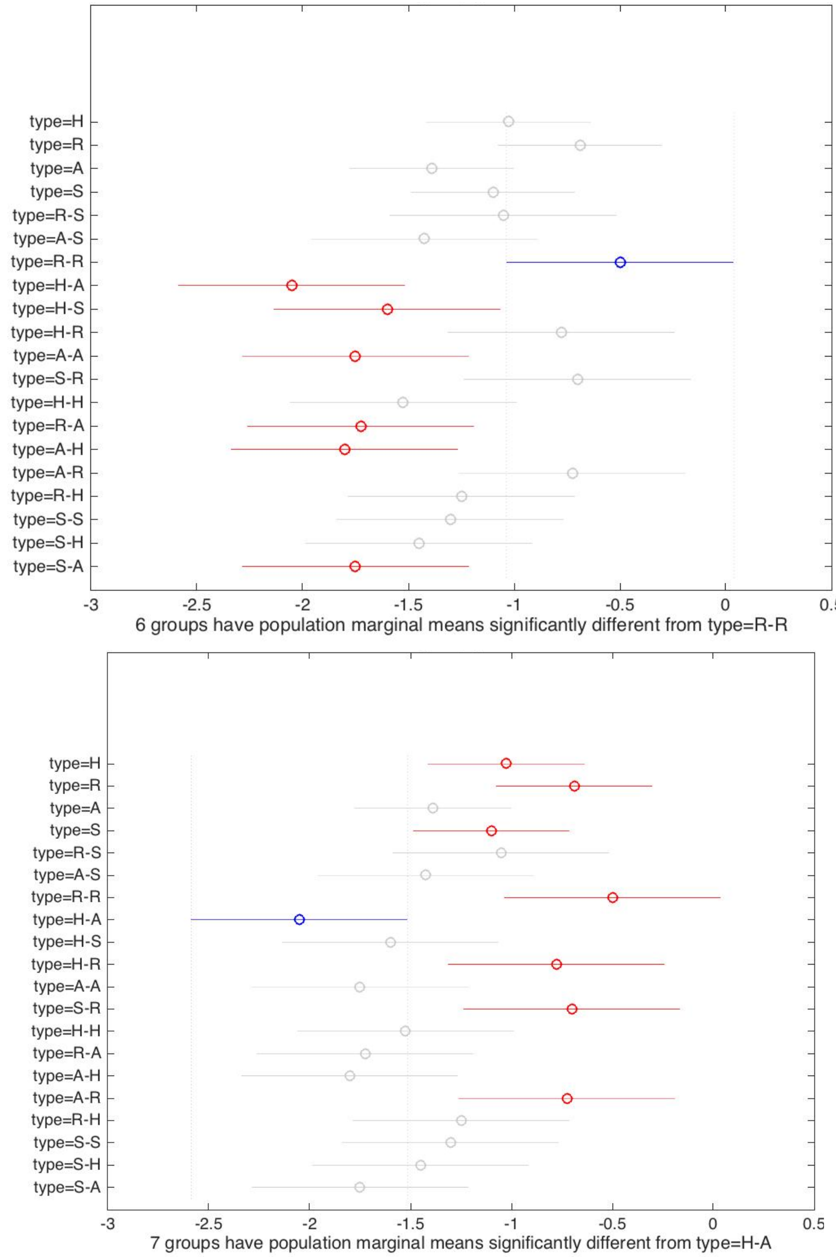


Figure 6.6: Multiple comparisons (with Bonferroni's correction). Haptic – Sound pairs represented. A) Relaxed-Relaxed has higher valence than all the combination with Angry sound, Haptic-Sad and Angry-Happy. B) Happy-Angry has lower valence than all but the Anger single modes, and all the combination with Relaxed sound.

To further disambiguate which single-mode stimuli had actual differences among their individual emotional types, we performed a one-way ANOVA for vibrotactile $F(1,3)=2.07$, $p=0.106$, which rendered non-significant. In the case of sound we observed a significant difference $F(1,3)=2.9$, $p=0.037$.

Figure 6.7 shows a multiple comparisons chart with Bonferroni correction (p=0.0125), where it can be seen that the sounds that have statistically significant differences t(78)=-2.88, p=0.0052 are Anger (low valence, high arousal) M=-2.24, SD=1.89 and Relaxed (high valence, low arousal) M=-1.06, SD=1.8.

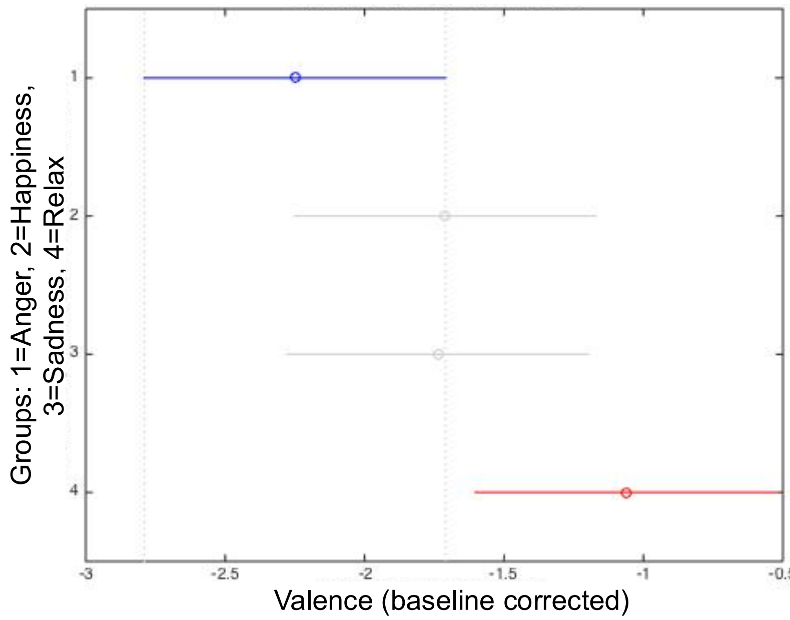


Figure 6.7: Multiple comparisons (with Bonferroni's correction). Only Anger and Relaxed present significantly different valences.

With this new information, it seems clear that despite not having statistically significant differences among haptic effects, the differences in sound were amplified with the haptic effect. Angry sounds got affected by pretty much every other haptic signal. Relaxed sounds remained with a higher valence and were not affected by any haptic stimuli.

With the results for Arousal and Valence we can conclude that H.1 holds in the case of Angry and Relaxed sounds, while H.2 holds true mainly for sounds and haptic signals that are opposing in the CMA quadrants.

6.5.2 Qualitative Analysis

Our qualitative assessment further support the findings from our quantitative analysis.

Emotional Expressivity: Overall, participants noted sound as being more effective in conveying emotions, while haptic stimuli were harder to interpret.

> *"With haptics alone (without sound), it is much harder to determine what the haptics are trying to convey. It is very open to interpretation because there*

are many possible meanings of what the haptics mean. Sound alone (without haptics) can convey more meaning because specific sounds are easier to associate with things we experience/hear in everyday life."

"I thought it was more difficult to relate an emotion to haptics. For some of the combinations, the emotion I was feeling was not listed."

"Haptics is good for alerts (why I usually associate it with stress), but sounds are better for conveying emotions."

Haptic stimuli were largely associated with alarms, stress, and high levels of arousal:

"Haptics seemed inconsiderate, obnoxious and annoying most of the time. They feel appropriate for alerting someone to pay attention, but they didn't really convey any emotions besides stress when they were alone."

"When the vibrations were on, they were distracting, and depending on the amount and speed of the vibrations my 'antsyness' went up. I became uneasy."

"The haptics tended to make me feel more energized, but I don't think they affected my emotions as much."

Furthermore, the length and intensity of the haptic stimulus influenced whether it was perceived as pleasant or not. Most participants enjoyed long and soft or slowly pulsing stimuli while sharp, abrupt, or strong stimuli were perceived as aggressive or negative:

"The sharp haptics was almost annoying. Like someone poking me. The more constant rolling vibration was a lot more calming."

"I liked a more sympathetic longer vibration. Like it was saying 'I am really sorry that I am waking you up, but you asked for it'."

Perceptual Dominance: With regard to dominance, slightly more participants described sound as the dominant stimulus. When asked about stimulus dominance, 15 participants (37.5 %) chose "sound. " The remaining participants were equally spread between "haptics" or "both," with 12 participants (30.0 %) each. While sound dominated in conveying the emotional tone, participants mentioned haptics as producing a greater emotional nuance.

"However, the combination of sound and haptics makes the most impact. The combination of haptics with sound can give different interpretations of what the sound is meant to convey."

"The sound was the dominant way I determined the emotion, it overpowered the haptics. The haptics were what I used to determine which category out of

the few that the sound could be related to. So for a sound that I felt could be fear, stress, or anger, the haptics would dictate which of those I placed it in."

"A soft haptic vibration soothed the sound of hard laughter (which could sound very annoying/intrusive). A pleasant sound coupled with an urgent, throbbing vibration made me angry (took away from what I wanted to feel). An urgent sounds with urgent/fast vibration made me feel like I wanted to flee rather than figure out what was wrong."

Combined Effect: The combination of stimuli was described to be enjoyable only when each haptic and sound stimuli were perceived to be matching in emotion. When the stimuli did not match, the combined trials were described as stressful.

"..., I would be inclined to say that the dominant frequency of the sound and the dominant frequency of the haptics need to be aligned. The misaligned sounds-haptics gave me the impression of bad design."

"There were combinations of sounds and haptics that angered or surprised me and I think it was due to a mismatch in the pairing. The laughter sound, when paired with the right vibration evoked happiness but when paired with another vibration made me feel angry or stressed."

"I expressed surprise at the early combinations of soothing/sympathetic music with vibration alerts, as they didn't seem properly paired... I didn't feel there was any emotional choice to match an alert or alarm..."

6.5.3 Usability Assessment

We investigated four usability metrics:

Potential Frequency of Use (Daily, Weekly, Monthly and Other)
Likelihood of Usage (1: least likely to 5: most likely)
Preferred Modality (top-down best to worst rank)
Perceived Dominance (top- down best to worst rank).

A descriptive summary of the data is presented in Figure 6.8. The radius axes were all normalized (from 0 to 1) to create a graph that helps visualize the highest values or rankings outwards from the center of the radius plot. 51% (21 out of 40) participants preferred sound. Likelihood of use for sound (0.82) was also higher than haptic (0.42) and combined stimuli (0.58). Potential frequency of use and perceived dominance were similar for all modes.

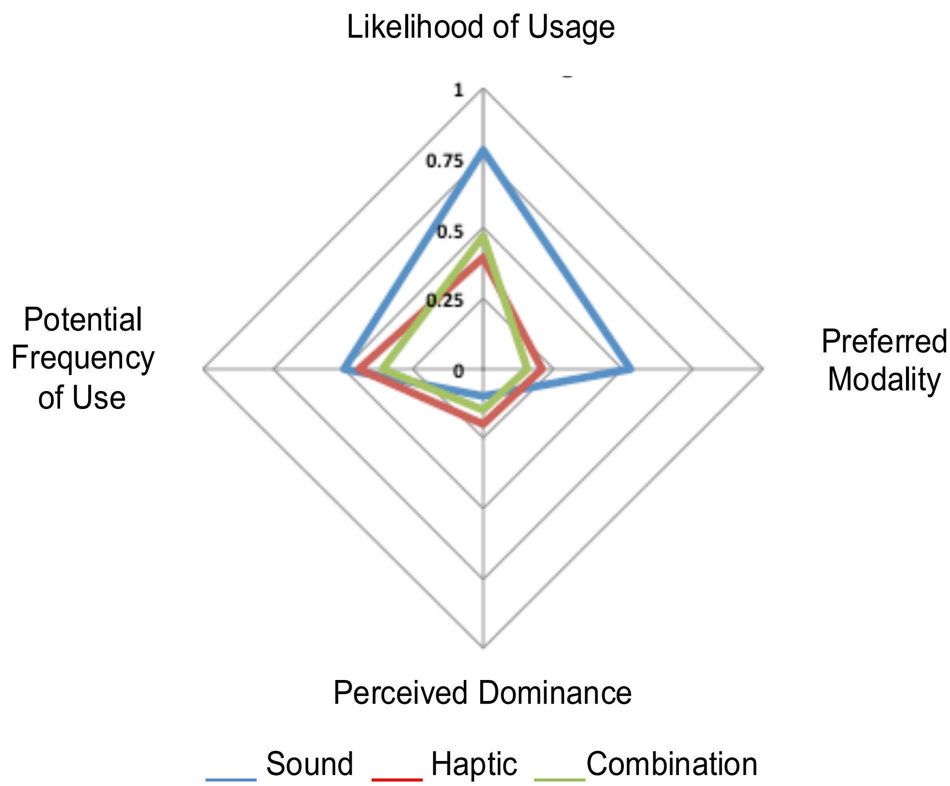


Figure 6.8: Usability metrics for Sound, Haptic and Combination (Sound + Haptic) conditions.

Gender Bias: Previous research showcases differences associated with gender and the perception of emotions [57]. We explored if there is a gender bias on usability. 21 females and 19 males participated in the study. We found that only sound had a gender bias with respect to its likelihood of use (chi-squared = 44.8509, df = 12, p<0.001) and preference (chi- squared = 22, df= 2, p<0.001).

In choosing which stimuli was more dominant in a multimodal condition, 8/19 (42%) males and 4/21 (19%) females chose haptics as being dominant; as opposed to 4/19 (21%) males and 12/21 (57%) females that chose sound. Males cited the alarming nature of haptics, and thus seemed to find haptics to be dominant over sounds even despite conceding to the greater emotional expressivity of sounds.

> *"I found it difficult to have a non alarming haptic - perhaps the lowest vibration level, longest pulse was closest (and least like any mobile phone alert-style vibration). Sound - even with the 3 or 4 limited clips used - allowed a richer range of emotions to be expressed."* [NOTE: Subject chose "haptics" when answering which stimulus was more dominant when combined.]

Females similarly noted the alerting nature of haptics, but would still find sounds to dominate over haptic stimuli.

> *"Haptics is good for alerts (why I usually associate it with stress), but sounds are better for conveying emotions."* [NOTE: Subject chose "sound" when answering which stimulus was more dominant when combined.]

Multimodal Communication: Table 6.6 shows the preferred modes for a communication application, which indicates that users overall prefer to use messages with sound only or with the combined multimodal method.

	Sound	**Haptic**	**Combined**	**None**
Counts	28	7	29	4
%	47,46	11,86	49,15	6,78

Table 6.6: Preferred communication modalities.

6.6 IMPLICATIONS FOR DESIGN

By looking at the quantitative results in isolation, we observe a couple of differences. Sounds do have better expressivity, however haptics is indeed a modulator. Sounds with negative feelings seem to have their effect enhanced with haptics. Furthermore, mixing opposed sounds and haptic stimuli generate a reduced Valence. We discuss our quantitative findings in light of the qualitative remarks to discuss the elements of design of multimodal (haptic and auditory) wrist-worn wearable devices:

Sound Stimuli Selection

Differences were present mainly between Relaxed and Angry sounds for both Valence and Arousal. This, however, contrasts with the qualitative remarks, where people highlighted the higher expressivity of sound. An appropriate design should consider personal preferences for sounds that elicit differentiable arousal and valence ranges. It can be argued as well that a dimensional view of affect is the only way to select different stimuli or to study its effects. A complement with a discrete view of emotions (Ekman's universal emotions [36]) could be of help.

Haptic Stimuli Selection

In the case of haptic stimuli we did not observed marked affective differences in the stimuli selection. We based our design on mimicking anthropomorphic touch using certain vibrotactile primitives such as poking, stroking, caressing, etc., which did not render proper affective responses. Perhaps a preliminary design challenge is to create an ontology of haptic stimuli with affective labels, where haptic primitives such as activation, decay, frequency, intensity, and duration, are used to describe affective primitives such as feelings, sensations and emotions.

Haptic as a Modulator of Sound

Overall, our findings show that despite a prevalence of sound as a better communicator of emotion, haptic stimuli play an important role in the overall affective experience. When combined with sound, a properly matching haptic stimulus can serve as a booster for the emotion expressed by sound. However, choosing a haptic stimulus that does not match the emotional content of the sound could reduce the emotional expression of the sound, by turning it into disgust or stress.

Matching Primitives

Beyond affective coherence, the design of multimodal stimuli should also consider the effects of matching certain stimulus primitives such as frequency, intensity, rhythm, etc. Properly matching these primitives may not increase the overall affective experience, however, any mismatch could generate an adverse emotional reaction.

Gender Matters

The observed gender bias showcases the importance to tailor sounds, haptic, and multimodal affective expressions to different genders, especially if the interest is specifically to evoke or communicate emotions. For males, in the case of multimodal stimuli (sound + haptics), haptic could play a stronger emotional modulation role.

6.7 FUTURE WORK

It is clear that more work needs to be performed to understand how devices can capture, but also express and help communicate emotions. One avenue is to focus on generating an ontology and a corresponding database of different haptic stimuli and the emotional responses that could potentially be generated by wearables and various Internet of Things devices. A complement to this analysis will be to understand other parts of the body where haptic stimuli could be applied. Finally, it would be relevant to investigate interactive scenarios where multimodal emotional expressivity can be of benefit, such as with autonomous vehicles, where mood and stress management will play an important role and where larger surfaces, such as the seat, could be used to improve the emotional expressivity of a haptic stimulus.

6.8 CONCLUSION

In the present study we have observed the potential affective expressivity of multimodal interaction of mixing haptic and auditory stimuli. We observed the lack of expressivity generated by the single-mode stimuli, despite existing theoretical foundations. We observed no statistically significant interactions, which contrasted with personal statements obtained through a qualitative assessment. Despite the lack of clarity from the quantitative analysis, it is observable that the interaction of the modes is not null, nor is dominated by one single-mode. Auditory stimuli tend to maintain certain dominance, but it is clearly not immune to the effects generated by the haptic stimuli. Finally, some design intuition is developed, which principally focused on contextual awareness, personalization, careful selection of single-mode stimulus, gender differences, and the potential need for mitigation of negative effects carried by mismatching stimuli.

7 Chapter 7
Urban Ambient Interventions

We examine how a smart LED lighting system that responds to positions and velocities of pedestrians can affect their emotions. We show that specific lighting patterns significantly increase positive affect among test subjects, and we propose a safety-to-pleasure hierarchy of interaction, in which users show more positive reactions to interactive lighting experiences once they have established a threshold of safety. Test subjects reported great interest and little concern about how their circulation data night be used, primarily because the collected data is anonymous (as opposed to cell phone logs) and because the generation of data and its display in situ built trust and transparency. Tested interactive lighting designs cut energy consumption by 75% while maintaining a positive affect similar to that of always-on lighting.

7.5 INTRODUCTION

Well-lit sidewalks can increase pedestrian's perceptions of pedestrian safety [143, 43] and security [42, 110], a.k.a. pedestrian reassurance, which is the confidence a pedestrian has to walk at night. Lights at night, besides improving visibility to avoid obstacles also improve perceptions of safety by allowing pedestrians to judge other people's emotions and intentions [50] and by helping them perceive escape or refuge that are relevant for perceived safety [49]. But such relevant tasks generate energy waste and light pollution, specially during late hours of the night and/or in places with low pedestrian traffic.

Can responsive lights that react to pedestrian movements improve pedestrian perceptions, while simultaneously reducing energy consumption and increasing the amount of data a city can collect about its most vulnerable residents, pedestrians at night? We wanted to explore pedestrian perceptions of safety and affect, perhaps while reducing energy consumption.

Expanding on the basic notion of a standard motion-sensitive light, which can deliver illumination "on-demand" with sophisticated arrays of networked motion sensors. We researched how temporal design factors for a linear array of lights along a sidewalk positively and negatively influenced pedestrian affect. Temporal factors addressed if the lights turn on before, during or after pedestrian motion, and also if the lights are activated gradually (slow) or instantly (fast). The lighting effects ranged from what one user described as a "*luminous path unfolding before [the user] with every step*" *to designs that* "*followed [the user] in a creepy way*". Spatial factors addressed lighting angles and the sizes of illuminated areas. Under controlled circumstances, these factors impact affect and that affect varies depending on a person's assessment of safety.

We conducted three separate experiments in the same setting, a dark underground hallway. First we evaluated design variations using a factorial design so we could define the affective impact of concrete parameters, such as activation time (timing) and transition speed. Second, we compared these interactive lights to an always on lighting condition. Our final experiment was designed to further expand our knowledge of the affective interaction design space for interactive lights.

We concluded our studies with queries of when and where interactive lights might be deployed to optimize positive feelings about locations while at the same time saving energy and providing a platform for residents to consider the impact of urban design on pedestrian flow and density.

7.6 **PREVIOUS WORK**

7.6.1 Urban Interactive Light Design

Haans and de Kort [49] describe the association between illumination and perceived safety (outlook, escape, and refuge). Although counter-intuitive, increased perceived safety happens when there is more light close to the pedestrian, rather than when most of the light is observed farther down the walking path. Fotios, et. al. [42] describe a disconnect between the optimal luminance level to improve perceived safety (10 lux), which is taken from a series of studies on the ratio between day and night perceived safety, and the actual illumination levels, which in most streets only fulfills the need to support obstacle detection (2 lux). The challenge is clearly how to find ways to deliver the appropriate level of luminance without a huge toll on energy consumption.

A series of workshops presented at DIS 2012 [3], CHI 2013 [4] and NORDICHI 2014 [2] have introduced several urban interaction design challenges such as the use of urban lighting systems in streets, city parks, and playgrounds; adaptation of existing UI paradigms to lighting systems; intelligent street lighting, etc.

7.6.2 Light Interaction and Expressivity

Ofermars, et al. describe similar themes in interactions with everyday lighting [109]. User interface and user context help define key elements of interactivity: motivation, dynamic lighting usage, context and routines, lighting system degrees of freedom, control availability, autonomous behavior and interaction qualities.

In 2012, Harrison, et al. researched how simple indicator light ON/OFF patterns communicated specific information to users [51]. The authors identified eight patterns strongly associated with information such as "turn on", "notification" and "low energy".

Our outdoor lighting system unites these interaction and information considerations to create a highly responsive sidewalk experience, and we proposed that the system will generate user expectations and needs comparable to expectations and needs for indoor lighting.

7.6.3 Implicit Interaction and Approachability

The theory of implicit interaction [65] proposes two key elements for implicit interactions: (a)dynamic or adaptive to the behavior and responses and (b) demonstrative, which implies expressivity. These elements informed designs of welcoming interactive devices [64]. We applied both concepts to street lighting to enable safe, welcoming and playful lighting interactions.

7.7 LIGHTING SYSTEM DESIGN

We designed and built a modular lighting platform that would allow the construction of arbitrary length linear arrays of interactive lights that can be programmed to respond to pedestrian occupancy and speed and direction of movement. After teaching a course in urban lighting at our institution, we learned that the key design parameters for outdoor lights and sensors are: *anti-theft, weatherization, and low cost*. The choice of sensors and lights was driven by these requirements. We chose low-cost passive infrared (PIR) sensors covered with a simple tubular 3D printed case (to reduce their sensitivity angle to its minimum) and low- cost weather resistant LED RGB lamps in order to make it easy to maintain and replace lost units.

Our system provides surface-covering sidewalk illumination. tvlight.com [142] has shown a system mainly designed for automobile interaction. advance the research in two aspects: (a) we show the positive affective impact of anticipatory lighting over reactive lighting and (b) we track pedestrian speed and direction instead of position.

7.8 METHODOLOGY

We introduce A non-traditional design methodology leveraging scientific approaches. A traditional "need-finding" approach would not have yielded all the conditions, which our experimental matrix method tested. The study set up works quite well for teasing out various design parameters that are useful when setting up a system like this in a real-life setting

7.8.1 Location

To learn about the impact of interactive lights on pedestrians, we set up a controlled experiment inside our building. We chose this location so that we could have a fully controlled environment to control for environmental confounds such as weather, illumination, noise, and other people. We chose a lab setup Color was a balanced RGB mix. Figure 7.1 shows the experimental setup with a visual and dimensions of the site and metrics of luminance and the light projections. Three different experiments address the following overarching questions: 1) What is the impact of activation timing and speed in the user’s affect and the energy profile of these interactions; 2) What are the differences in affect and energy consumption between interactive and always-on systems; and 3) What are additional timing and ramp design considerations for optimal affective and energy outcomes.

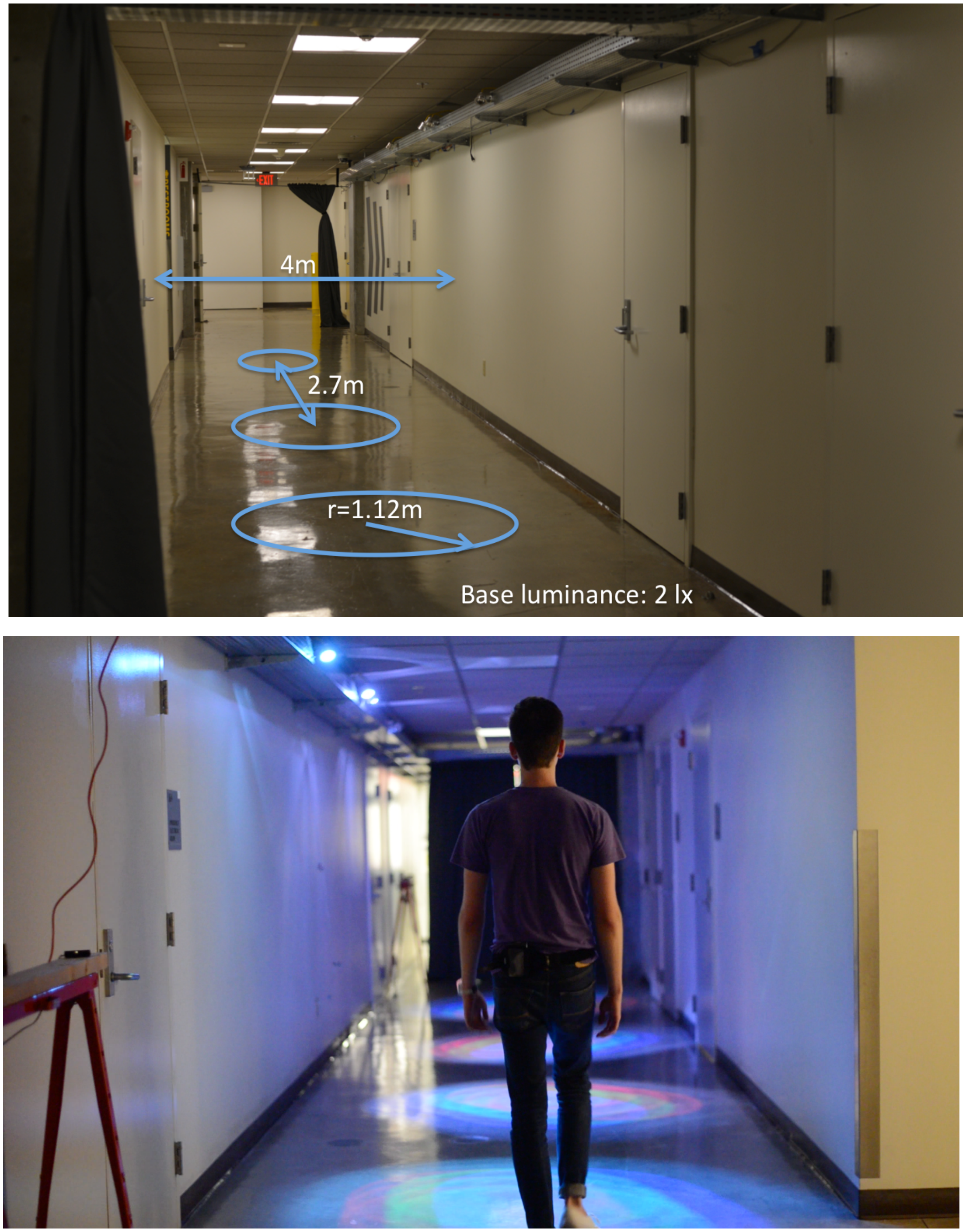


Figure 7.1: Experimental setup a) metrics: light radius = 1.12m, distance between lights = 2.7m, hallway width = 4m, hallway length = 30ft, base luminance with lights off = 2lx. Note: low illumination not displayed in this image; b) User walking with all lights on. Max luminance per light = 40 lx. Note: Image captured with high sensitivity camera, appears to be brighter than reality.

7.8.2 Method

For all three experiments, participants were recruited from our institution and the surrounding community, varying in age from 18 to 70 years old. Participants received $20 as compensation and were screened for any physical adverse reactions to light exposure. Each of the three experiments had three stages: (a) initial questions to determine baseline affect, (b) walking up and down a hallway that was solely illuminated by the interactive lights, and (c) post- experiment questions.

In *Stage A* participants rated their affect, watched a relaxing video, and then rated their affect again. This helped us determine baseline affect.

In *Stage B* participants were exposed to light conditions made up by *Timing* and *Ramp* activation factors. By *Timing*, we mean when the light comes on as the user walks by. The timings are: *Before* (the light comes on ahead of the user), *During* (the light comes on when the user gets within range of the light—like a spotlight), *After* (the light comes on behind the user), and *Random* (the lights turn on and of randomly as the user walks by). By Ramp, we mean how the light comes on—it either fades in slowly (we call this *Slow* transition), or comes on abruptly (we call this *Hard*). The conditions were combined to create the different conditions. E.g. When the user walked down the hallway with lights fading in before them (like a rolling carpet) they were exposed to the *Before + Slow* condition. If the lights faded in, but came on behind them, this would be the *After + Slow* condition. There was one condition where there were no lights at all and one where lights were *Always On*. Figure 7.2 shows a representation of the *Before + Fast*, *During + Fast* and *After + Fast* depictions of the lights, as they appear projected on the floor while the participant walked.

The participants were exposed to every combination twice and then were asked to answer three questions related to the Circumplex Model of Affect (CMA) [127]: a) “What is your current level of energy? (from tired to excited) - Arousal, b) “How pleasant are your feelings right now? (from unpleasant to pleasant) – Valence and c) “What is your current level of stress?” (from low to high). We tested twice in order to reduce novelty effects. Figure 7.3 shows the average of the answers to the Energy/Arousal and Pleasantness/Valence questions for Experiment 1 mapped into the CMA.

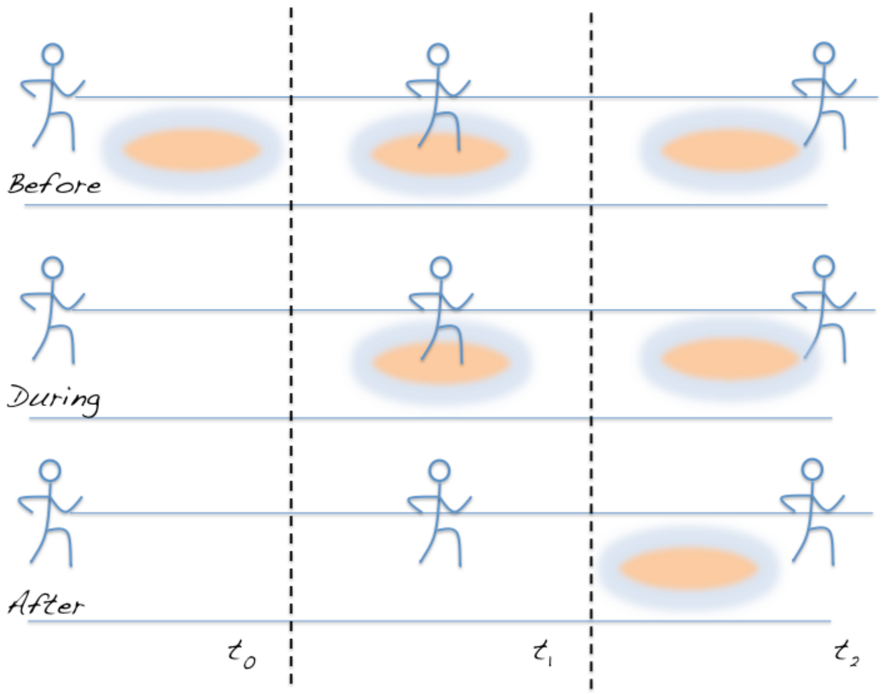


Figure 7.2: Fast Transition Lighting Interactions

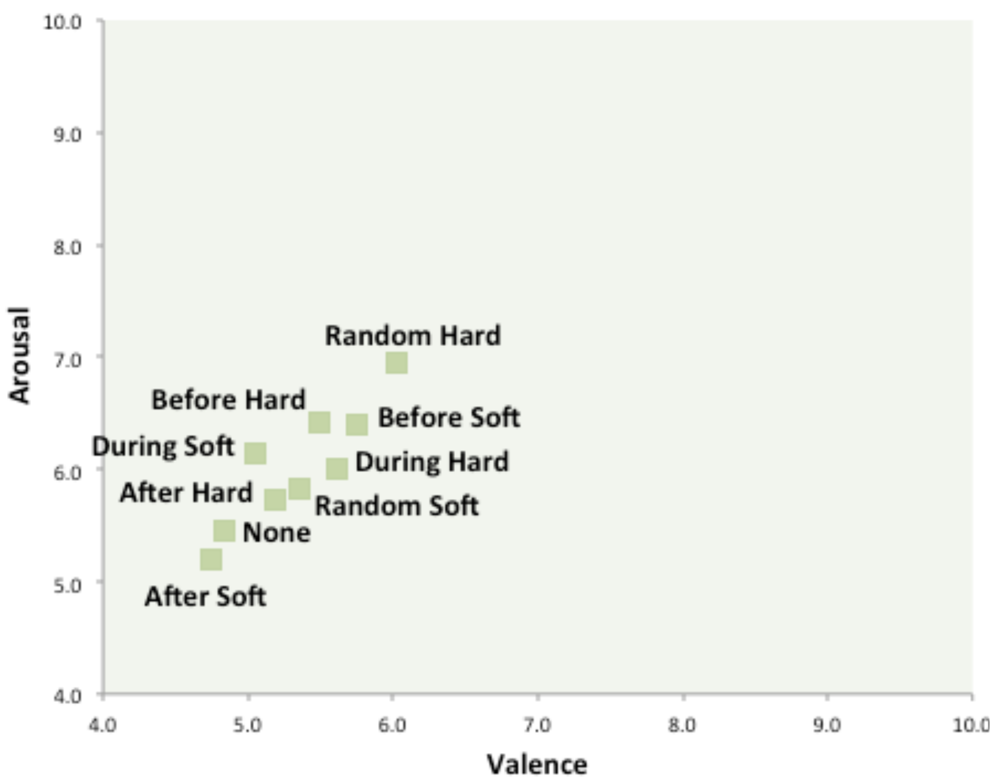


Figure 7.3: Circumplex Model of Affect (CMA) mapping of the Arousal and Valence metrics for each condition in study number 1.

In *Stage C* participants completed two parts: a *survey* and a *card sorting* exercise. The survey was composed of two parts as follows:

1) Immediately after the experience stage, users were asked for the conceptual model. Users were asked to describe in their own words the experience they just had with lights.
2) After revealing to the users the two factors (Timing and Ramp) and the order in which the conditions were presented, the user was asked to describe: their emotions, their preferences, and the type of use their would give to these lights in urban scenarios.

For the card sorting activity, we gave the users 16 different cards depicting: *crosswalk, tunnel, overpass, construction site, metro, airport conveyors, museum, snowy street, pleasant street, nature, backyard, battle ship, war zone, industrial bridge, alley, prison*. We asked people to rate their emotions with regards to the different places from dusk to dawn and to fill up a board where they placed the cards in the preferred light category—e.g., if a user really liked the before-hard condition we would see many cards in that category. We used cards to determine where users would like to see the interactive lights they just experienced.

7.9 EXPERIMENT 1: INTERACTIVE LIGHT FACTORS

7.9.1 Study Design

Objective

In this study we compared interactive modes that were responsive to the speed and position of the user.

Experiment Design

We chose a factorial (2x4) within-subjects design with N=36 participants, 19 females and 17 males with a mean age of 29.7 years. 38.9% of them were students and most of the rest were employed with a variety of trades.

We measured Valence, Affect and Stress. We manipulated Transition (*Slow* or *Fast*) and Timing (*Before*, *During*, *After*). We had two controls: *Random* and None.

Hypothesis

H1.1 – Activation has main effect on Affect

H1.1.1 – Timing has a main effect on Affect

H1.1.2 – Ramp has a main effect on Affect

H1.1.3 - Timing has an interaction effect with Ramp
H1.2 - Before is more popular than other Timing conditions.
H1.3 – A Slow transition is more popular than Hard
H1.4 – Activation (Timing and Ramp) has an interaction effect with the feelings associated to a specific urban place

7.9.2 Quantitative Analysis

We performed a within-subjects two-way ANOVA for each dependent variable.

Arousal

No interaction or main effects were significant for Arousal.

Valence

Interaction effects between Timing and Ramp were not statistically significant $F(3,36)=1.43$, $p=0.3253$. Ramp had no effect, while Timing had an effect on Valence. $F(3,36) = 3.74$, $p<0.05$. Multiple comparisons (Bonferroni) showed differences between *After* ($M=-1.5417$, $SD=1.9205$), *Random* ($M=-0.6111$, $SE=2.1$), $p<0.05$ and *Before* ($M=-0.5972$, $SD=1.9763$), $p<0.05$. Therefore, we reject hypothesis *H1.1.2* but we accept *H1.1.1* only for Valence. This implies that Timing (*Before*, *During*, *After*) has an effect on pleasantness.

Energy

The amount of energy consumed per condition measured in Watts*hour (Wh) is a function of all the time the lights were on during the trajectory of the experiment (30 ft). Estimations for energy consumption were calculated based on the average walking speed from our participants of 1.21m/s., and an occupancy of 1 pedestrian per every 30 feet. As it can be observed in Figure 7.4, the amount of energy depends mainly on the *Transition* function (*Fast* or *Slow*). *Random* functions do not depend on the velocity of the pedestrian, and respond to a 50% chance function, so its consumption of energy is higher.

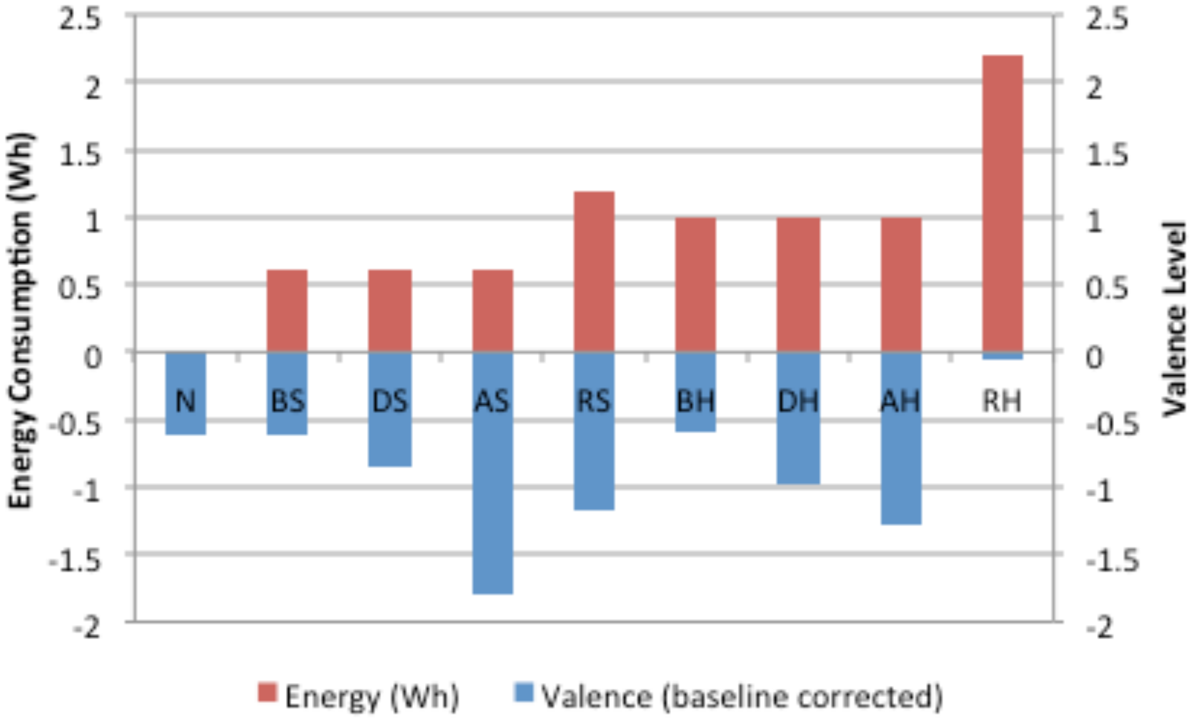


Figure 7.4: Energy consumption (Wh) in red, contrasted with Valence (baseline corrected) in blue for each light condition.

Having an always one base light or having more lights ahead of the user should increase these energy profiles - we discuss this issue in experiment 3. Timing does not affect the energy outcome. Of course there could be other implementations that could be more or less energy efficient. Figure 7.4 also shows the average values (baseline corrected) for Valence and Arousal per light condition. There is moderate correlation between energy consumption and Valence ($r=0.3786$).

Preference analysis

To evaluate preference (from 1 = most preferred to 5 = least preferred), we performed a Kruskal-Wallis non-parametric analysis of variance for each lighting factor. Timing conditions were found to be different $H(4)=86.79$, $p<0.0001$. Post-hoc comparisons (Bonferroni) revealed that *Before* (Median=1, SD=0.5248) was preferred to all other conditions. *During* (Median=2, SD=1.079) was better ranked than *After* (Median=4, SD=0.8669) and *None* (Median=5, SD=1.331). *Random* (Median=3, SD=1.1557), *After* and *None* were not different from each other. No differences were found between the Ramp (*Soft* and *Hard*) functions. Hypothesis H1.2 holds and but H1.3 does not. These results imply that again, Timing influenced participant’s preference.

7.9.3 Post-test Qualitative Analysis

All participants (N=36) responded to a post-test survey. Key results extracted from coding the open-text questions are summarized below.

Conceptual Model: People differentiated between lights ahead of them (*Before*), lights that turned on as they passed (*During*) and lights that turned on behind them. About 20% of the people were able to perceive and describe all the interactive light factors, while virtually all users (98%) detected at least the *Before* choice and another choice other than *None*. 15% of the people associated the patterns with their position.

> *The first pattern was that the light turned on in front of me while I walked in the dark. / The 2nd was that the light turned on after me when I walked in the dark. / The 3rd was the colorful light turned on to light my way when I walked in the dark. / The 4th was that the colorful light followed my steps when I walked in the dark. / The 5th was there was no light at all.*

Best interactions: 41.7% of the people selected *BS* as the best interaction. About 30% of the people thought the modes were useful. 25% of the people saw a *Slow* transition as more relaxing while 20% considered a *Fast* transition to be more useful, as it provides more visibility and is more predictable.

> *Before soft* [slow transition] *gave me leisure time to adapt to whatever was coming into view as I progressed. The softness of the increasing illumination as I approached was relaxing, appealing, pleasant.*

Worst Interactions: None, *AH*, *DH* and *RH* were the worst conditions. People found the hardness "harsh". An interesting nuance observed about *After* lights was that 30% of the people considered it very scary, creepy, threatening, as if they were being followed.

> *It* [After + Slow] *was very creepy and felt very threatening. It reminded me of someone coming up behind me at night and it just kept on happening as each light creeped on.*

Walking at Night: When asked if people would chose to walk more at night with interactive lights, 67% responded affirmatively. They believed that lights would make it safer, and the interactive nature would make it more comfortable with the added benefit of lower energy consumption and less light pollution.

> *While I like lights on all the time, it would be neat to have lights that only came on when needed, that way you could also see if someone was walking from a long distance away as lights would come on further down the street. Also, this would mean less light pollution.*

Those who would not walk more at night (33%) mostly believed that walking at night is inherently unsafe and that lights are not a major role towards improving safety.

> *Lighting does not change the variables that make walking alone at night dangerous. Predators will continue to lurk around regardless of the lights.*

7.9.4 Card Sorting Analysis

Before sorting the cards, users were asked to describe the places where they would use interactive lights. 31.6% of the votes were for Urban (non-residential), 22.4% were for Leisure, 14.5% were for Residential as well as for Traffic/Commute and 17.1% were for other various places.

After responding to a conceptual model of the

Figure 7.5 shows the affect board. People were first asked to rate the places based on their preliminary affect towards that place. More specifically, we asked them to imagine you are there at night, and please rate your feelings (Negative and Positive) and their excitement (Boring/Calm and Exciting).

Figure 7.6 shows the light selection board. People were asked to imagine they are at night in each place and to select the best light condition that could be used for each scenario. They were also told to explain their selections in open text form.

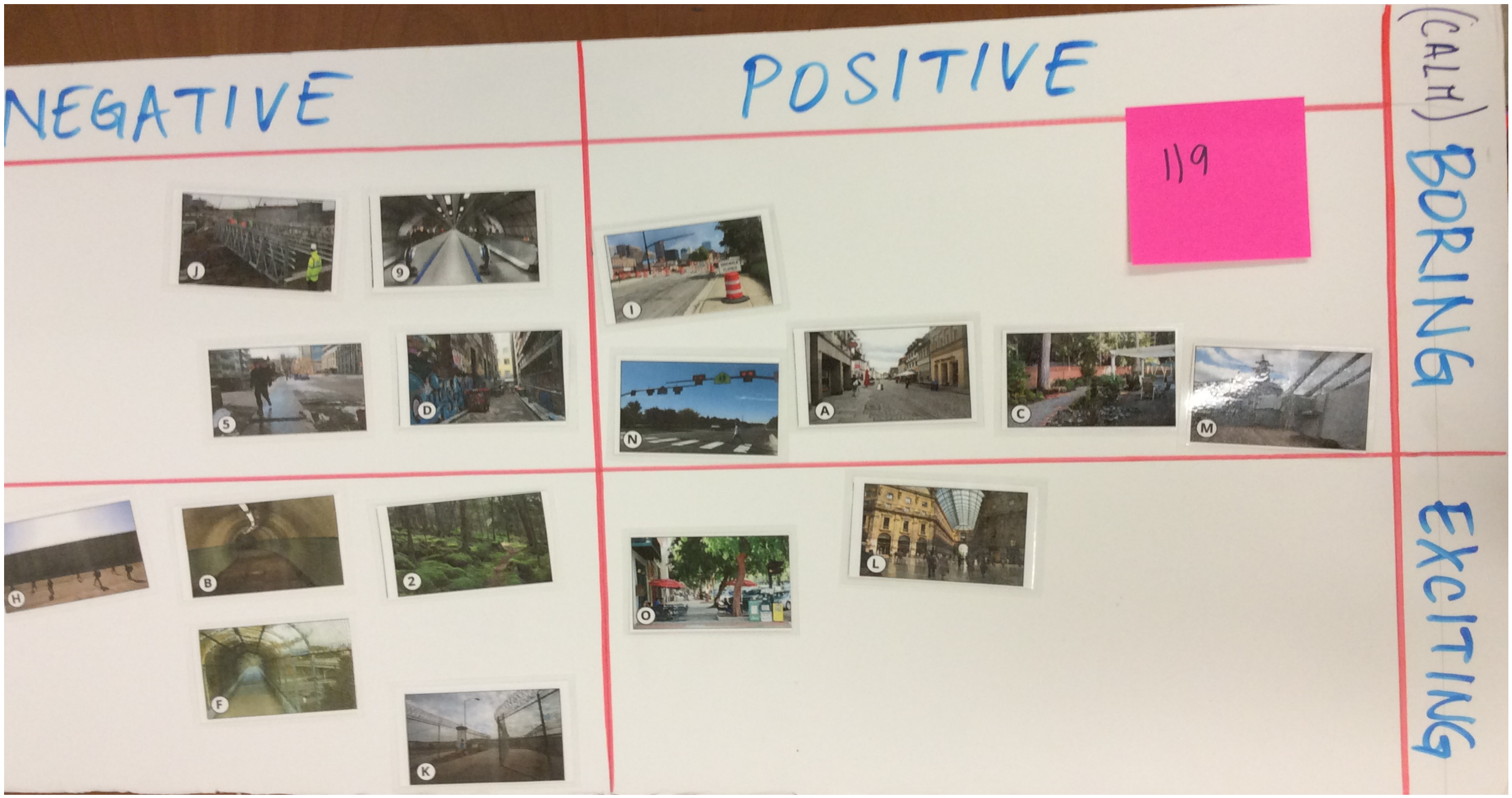


Figure 7.5: Board used to sort the different cards with places into the best affective classification.

Figure 7.6: Example of the light conditions board. Used to match places with light conditions.

Table 7.1 shows the distribution for the 16 places we used for the card sorting exercise rendered a balanced distribution across the positive/negative feelings (Valence) and the excitement level (Arousal).

Places	Negative Feelings	Positive Feelings	Total
Calm / Boring	137 (23.8%)	143 (24.8%)	**280 (48.6%)**
Exciting	170 (29.5%)	126 (21.9%)	**296 (51.4%)**
Total	**307 (53.3%)**	**269 (46.7%)**	**576**

Table 7.1: Counts of places with respect to the feelings and level of excitement.

A four-way ANOVA revealed that there is an interaction effect between the feelings towards a place and *Timing* $F(3,36)=8.107$, $p=0.0129$ and *Ramp* $F(1,36)=4.4177$, $p=0.022$. We therefore accept hypothesis *H1.4*. As observed in Figure 7.7a, *After* (1.4%) and *None* (5.7%) were not very popular options. *Random* (12.8%) was much more preferred for positive places (83.8%) than for negative places (16.2%). This means that there is an inverse relationship between the feelings associated with a place and the type of interaction. Interactions that are less "useful" are chosen mostly for positive places, while interactions that are considered more "predictable" are preferred for negative places.

> *The purpose of light is well, to light the way and so having light after you have walked does not light the way, Having light in nature disturbs nature. Having soft* [slow transition] *light in public places might be a nice touch, Having light at intersections might prevent pedestrian accidents if they and drivers can easily see, There should be light during a tunnel trip.*

As observed in Figure 7.7b, a *Fast* transition (57.3%) was slightly preferred than a *Slow* one (42.7%). People preferred a *Fast* transition for negative places (67.5%) versus positive ones (32.5%), while people preferred a *Slow* transition for positive places (63.8%) versus negative ones (36.2%). The data supports the notion of an inverse relationship between the preference to use more predictable, less startling interactions for negative places and interactive options for better places.

> *Large, unconfined outdoor spaces would be good to not have lights. pleasant, safe places would be nice to have lights come on softly while you are walking past. A backyard might be a good place to have lights come on softly after you pass. Tunnels, crosswalks and construction zones would be well suited to have lights come on hard and before you pass. Neutral urban places would be nice to have lights come on before and softly.*

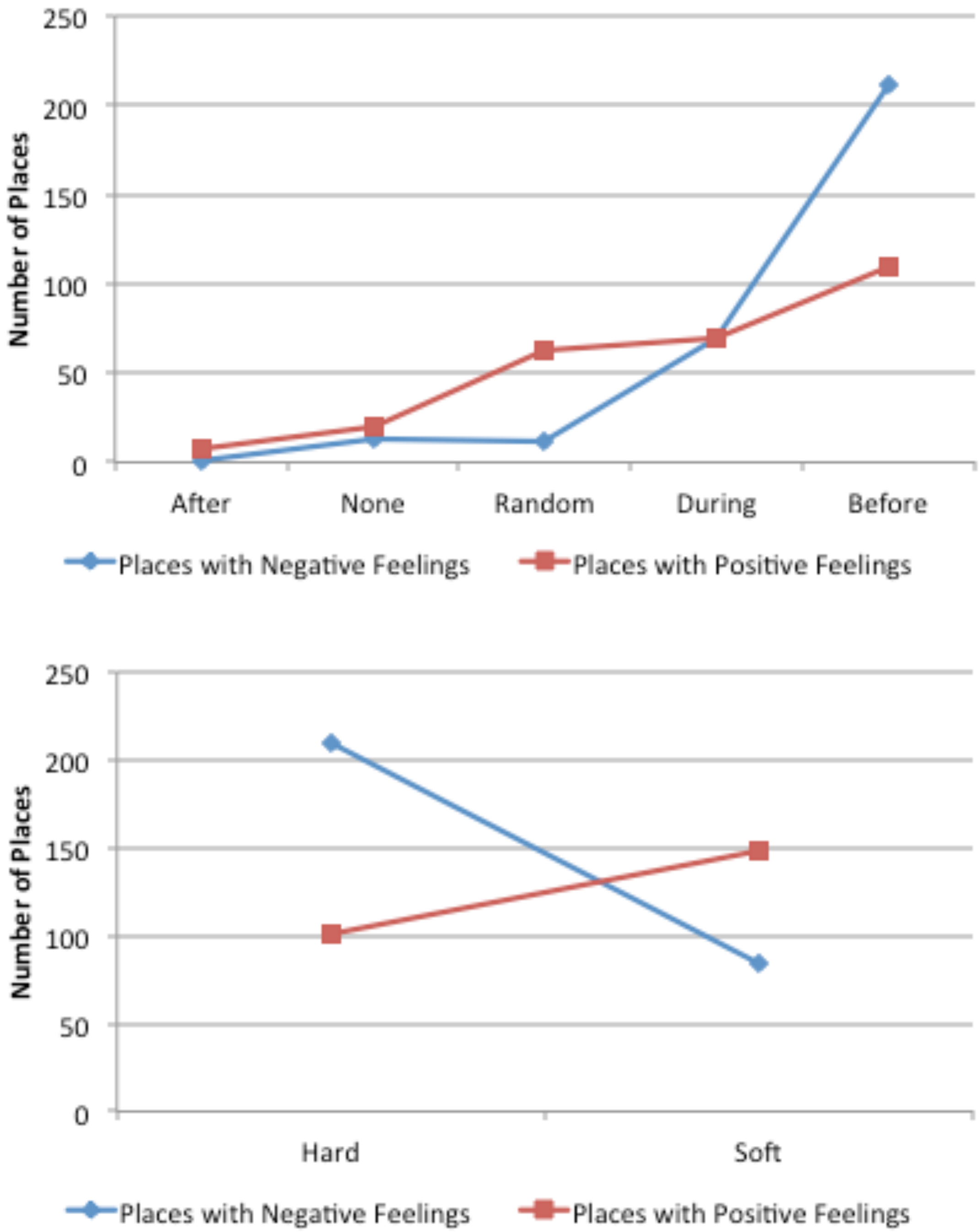


Figure 7.7: Interaction between Valence (feelings) about a place and (a) Timing and (b) Ramp factors.

7.10 EXPERIMENT 2: INTERACTION VERSUS ALWAYS ON

7.10.1 Study Design

Objective

In this experiment we want to compare the main interactive light modes with a static model that we call Always On.

Experiment Design

We chose a factorial (2x4) within-subjects design with N=30 participants, 15 females and 15 males with a mean age of 29.9 years. 43.3% of them were students and most of the rest were employed with a variety of trades. Transition had 2 levels: *Slow* and *Fast*. Timing, had 4 levels: *Before*, *During*, *After* and *Random*, plus a *No Light* control condition and an *Always On* static condition.

Hypothesis

H2.1 - Before + Slow has equal valence to Always ON
H2.2 – Interaction between Ramp/Timing for place feelings.
  H2.2.1 – Before is preferred to Always ON
  H2.2.2 – A Slow transition is preferred to Fast
H2.3 - Always ON is preferred for positive places

7.10.2 Quantitative Analysis

Causal Analysis

1-sample t-tests comparing the difference between *Always On* (M= -0.9667, SD=2.4842) and *Before* (M=-0.7, SD=1.779) were not statistically significant for Valence, t=-0.9322, p=0.352, nor for Arousal, t=-0.2356, p=0.8154. *Always On* falls in the same group as the *Before* and the *During* options, so we accept hypothesis *H2.1*.

Energy Analysis

As shown in Figure 7.8, despite a non-significant difference in Valence or Arousal, the energy expenditure for *Always On* is 3.5 to 7 times higher than any of the interactive modes. We spend a lot more energy without obtaining any gains. Correlation between Energy and Valence was actually weak (negative), r=-0.2553, between Energy and Arousal was weak (positive), r=0.2393.

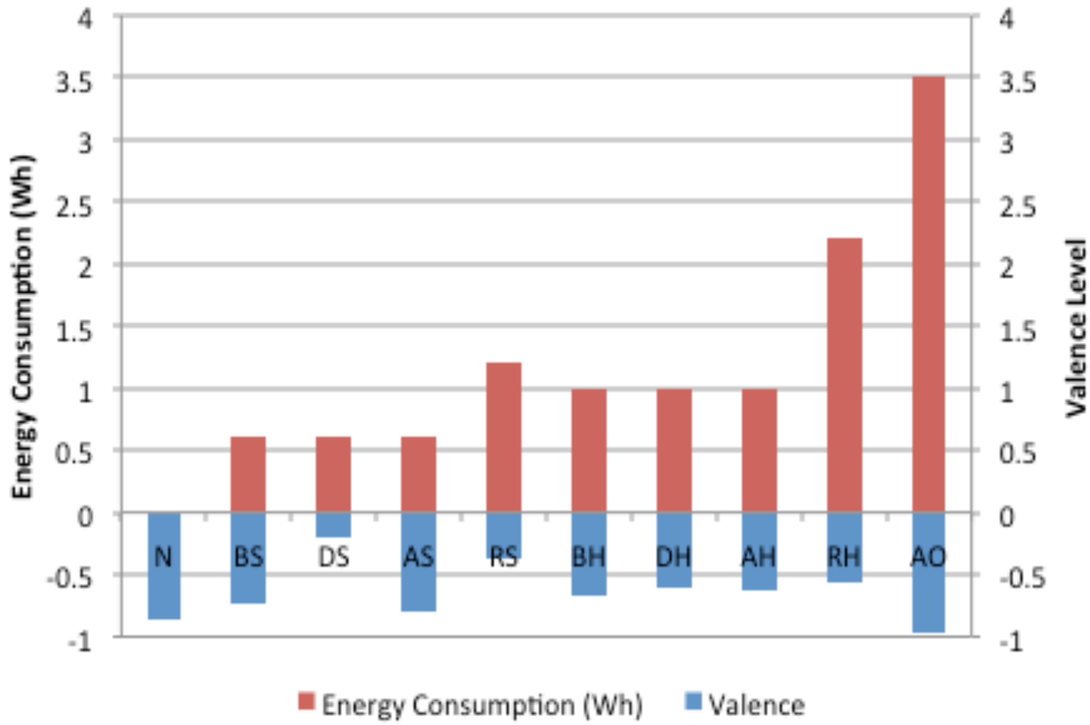


Figure 7.8: Energy consumption (Wh) –red- contrasted with Valence (baseline corrected) –blue- for each light condition

Finally, we measured the power consumed by the electronics involved to make the interactions. We wanted to compare an interactive system versus non interactive system. Energy drawn by the microprocessors and electronics does not surpass 7% the amount of energy spent by the Always On option.

Preference analysis

With regards to preference in *Timing* (1 = most preferred to 5 = least preferred) a left-side Wilcoxon non-parametric test indicated that the *Before* condition (median=1.5) condition was preferred as compared to the *Always On* condition (median=2), $Z=-2.1053$, $p<0.05$. We accept hypothesis *H2.3.1*

In turn, *Always On* was preferred as compared to the *During* condition (median=3), $Z=-1.7896$, $p=0.0368$. Preference for *Ramp* functions (1 = most preferred, and 2 = least preferred) tested with a Wilcoxon non-parametric test revealed that a *Slow* transition was preferred versus a *Fast* one, $Z=4.3935$, $p<0.001$. Hypothesis *H2.3.2* holds.

7.10.3 Post-test Qualitative Analysis

All participants (N=30) responded to a post-test survey. Many of the findings from the open-text question analysis are similar to the ones found in experiment 1. The main contrast was the preference for *Always ON* instead of *Before Hard*.

Walking at night: When asked if people would walk more at night 83% of people said that Interactive lights will entice them to walk more, while 77% people said that *Always On* would entice them to walk more. Those who responded yes to both modes revealed that any light condition was OK to walk, while some believed that Interactive modes were more vivid and felt like a company, as opposed to static which was more comfortable.

> *As long as there's lights in general, yes [I would walk more often at night]. That's always a plus. The interactive lights make it feel like there's another presence WITH you (in a comforting, secure way).*

When asked what percentage of time should be interactive or *Always On* modes be used, the decision is split. *Always On* (48%) was slightly less popular than of the time versus Interactive modes for 52%. Regardless of their main preference, most people considered *Always On* as more secure or predictable, while Interactive modes could save energy.

> *I think interactive lighting is preferable for comfort, for saving energy, and for enjoyment, but static light is necessary for safety reasons sometimes.*

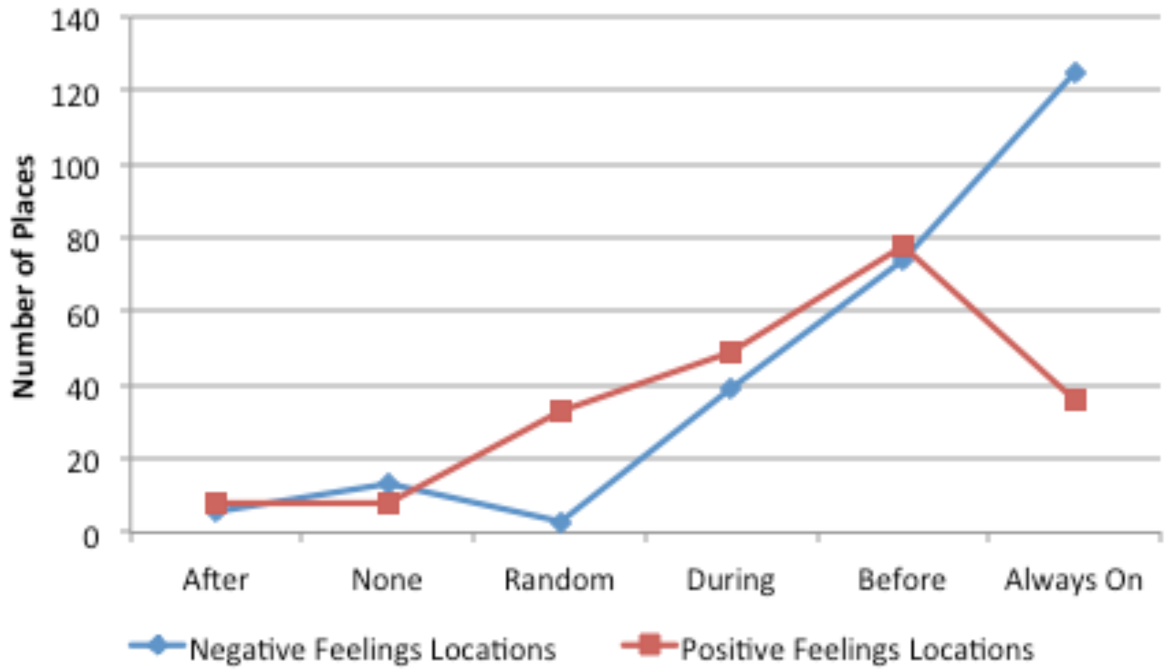
7.10.4 Card Sorting Analysis

The card sorting analysis revealed similar results between experiment 1 and experiment 2. It was good to observe again that the different places were distributed evenly across the difference conditions (see Table 7.2).

Places	Negative Feelings	Positive Feelings	**Total**
Calm / Boring	134 (28%.4%)	102 (21.62%)	**236 (50%)**
Exciting	126 (26.7%)	110 (23.3%	**236 (50%)**
Total	**260 (55.1%)**	**212 (44.9%)**	**472**

Table 7.2: Counts of Places with respect to the feelings and level of excitement.

A four-way ANOVA revealed that there is an interaction effect between the feelings towards a place with *Timing* $F(4,36)=235.844$, $p=0.0012$ (Figure 7.9a), as well as between feelings and the *Ramp* function $F(1,36)$, $p=0.0013$ (Figure 7.9b).



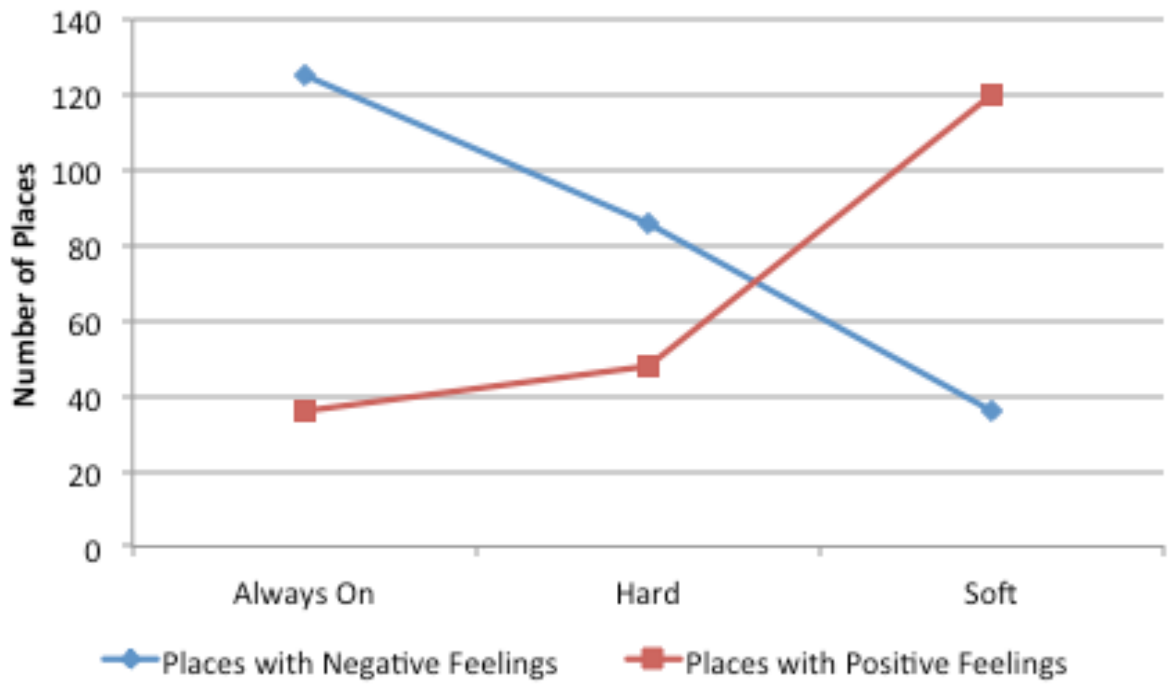


Figure 7.9: Interaction between Valence (feelings) about a place and (a) Timing and (b) Ramp factors

Hypothesis *H2.2* holds. It is interesting to observe that the *Always On* condition was mainly chosen for negative places. We accept *H2.4.* People explained that *Always On* was useful for dangerous places, while they preferred the *Before + Slow* option for not dangerous ones.

> *I put the places where I thought would be the least safe under the always on category. Places where I would look forward to seeing something ahead I put in the before (soft) category. Places where I think I would be around people just walking around, I put in the during (soft) category.*

7.11 EXPERIMENT 3: BEFORE + SLOW PRIMITIVES

7.11.1 Study Design

Objective

This experiment wants to test additional factors that may play a role in the design of a *Before + Slow* interaction.

Experiment Design

We chose a full-factorial within-subjects design with N=28 participants, 14 females and 14 males with a mean age of 27.9 years. 64.3% of them were students and most of the rest were employed with a variety of trades.

We measured Valence, Arousal and Stress. We also measured Fun and Usefulness. We manipulated the transition function this time to make it either *Smooth* or to have a *Ripple* effect. We chose this difference to observe if there is value in having a more noticeable change in (soft) transition effect. We modified the *Look Ahead Horizon* to make it either *1-light* (i.e. the same as the condition in the past 2 experiments) or *2-Lights*. We wanted to test if more light ahead of you had additional value. Finally, we added a *Base Light* (*Yes / No*) factor to see if having a dim (5% intensity) base light had additional value. From now on we will call each combination according to their initials, e.g. *1LNBR = 1-Light + No Base Light + Ripple*, or *2LBS = 2-Lights + Base Light + Slow*. Figure 7.10 shows the DVs to mapped to the upper quadrant of the CMA.

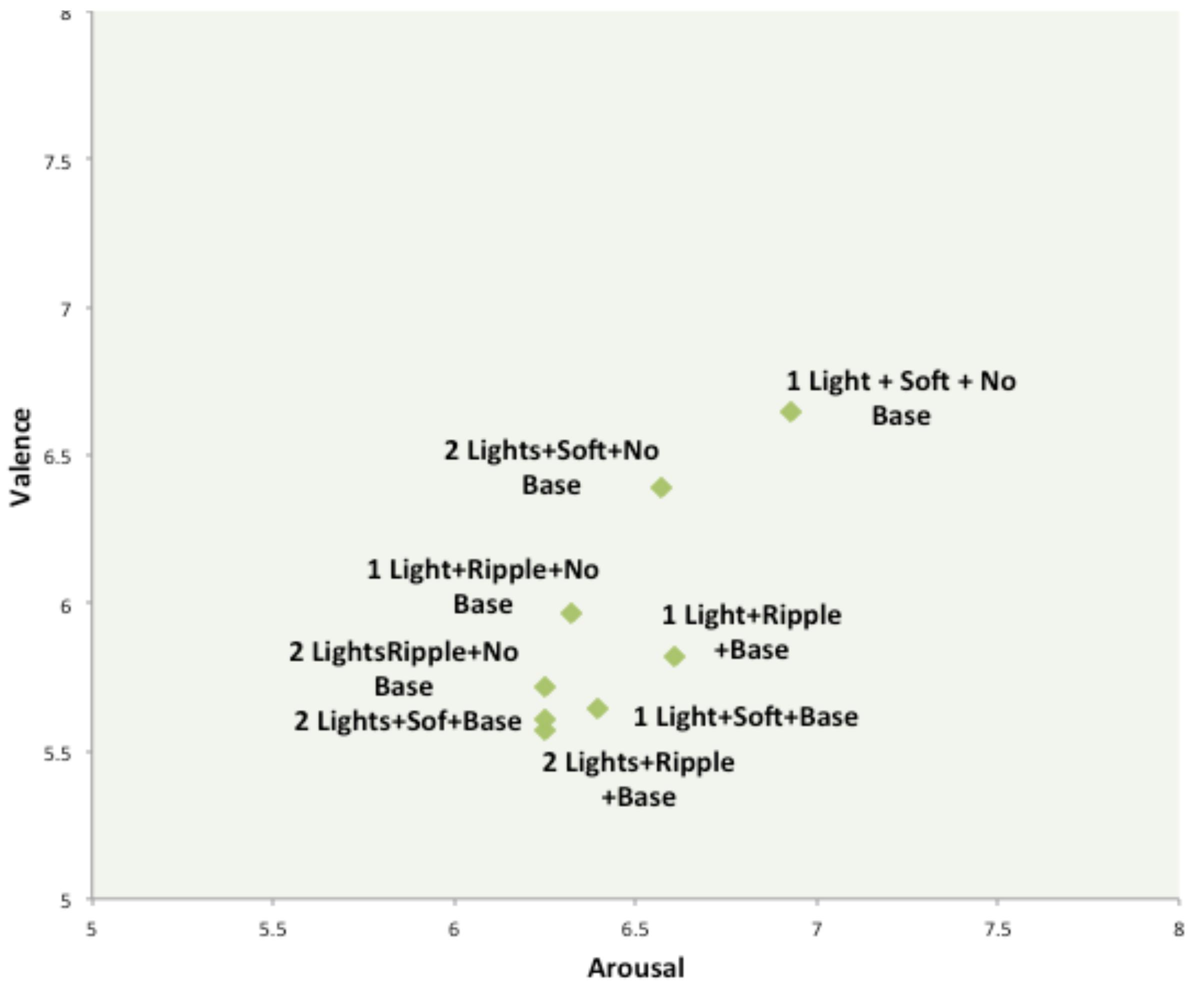


Figure 7.10: Interaction between Valence (feelings) about a place and (a) Timing and (b) Ramp factors.

Hypothesis

H3.1 –Ripple, Look Ahead, and Base Light have a main effect in Arousal and Valence
H3.2 – There are pairwise interaction effects between DVs
H3.3 - Preferences
- *H3.3.1 – 2-Lights are more popular than 1-Light*
- *H3.3.2 – Smooth Transition is more popular than Ripple*
- *H3.3.3 – Base Light is more popular than No Base Light*

7.11.2 Quantitative Analysis

We performed a three-way within-subjects ANOVA. We discovered no interaction effects between the manipulated factors and the measured variables. We found some main effects described below. Hypothesis *H3.2* holds.

Valence

We found *Ripple* to have an effect $F(1,27)=11.37$, $p<0.01$. Number of lights had significant effect $F(1,27)=5.96$, $p<0.05$. Some people described the feelings associated with *Ripple* as annoying. Hypothesis *H3.1* holds only for *Ripple*.

> *The flickering thing made me feel uneasy. As if something I was not expecting could suddenly happen.*

However, as explained in the qualitative analysis section, some people found Ripple to have some entertainment or even utilitarian value.

Arousal

We found an effect of the *Look Ahead* function on Arousal $F(1,27)=4.67$, $p<0.05$. People described the *1-Light* ahead horizon as more surprising than the *2-Lights* ahead horizon, which they found more relaxing and predictable. Hypothesis *H3.1* holds only for *Look Ahead*.

> *I liked being able to see where I was going before I got there. The 2 lights w/ base lights w/o ripple did just that...*

Energy Analysis

In terms of energy, the extra amount required to provide base light and two-lights of illumination in front of people was about twice the original (*Before + Slow*) option. However, even though we light up twice as much, this option still represents only 55% of the energy consumption generated by the Always On option from experiment 2.

Preference Analysis

We performed three pairwise Wilcoxon rank-sum tests to compare across the different factors. *Smooth* (median=1) is more preferred than *Ripple* (Median=2), $Z=-4.7581$, $p<0.0001$. 2-Lights (Median=1) are more preferred than *1-Light* (Median=2) $Z=-3.6986$, $p<0.001$. *Base Light* (Median=1) is preferred to *No Base Light* (Median=2), $Z=-2.6392$, $p<0.01$. Therefore, we hypothesis *H3.3* holds. This means that people prefer to have a base light, more light ahead of them and a simple ramp activation function to turn-on the lights.

7.11.3 Post-test Qualitative-analysis

All participants (N=28) filled out a post-test survey. The subtlety of these interactions generated some confusion in the way people saw differences between lights. Some people focused on the colors, others on the intensity, and others on the timing. Some people found the "flickering" (*Ripple*) disconcerting (but not scary), but they stated that this could be perceived as "fun" under certain circumstances. People preferred *2-Lights* ahead with a Base light, as they found this provided the best visibility.

> *The light patterns which flickered on made it difficult to keep a normal pace because I was not able to see what was in front of me. The lights that came on immediately (there were a couple of these) allowed me to keep a good pace throughout the whole hallway. There was one lighting pattern that started extremely dim and one by one would flicker on in front of me reminded me of lights that would be in a dark alley way. I quite enjoyed the lighting pattern that immediately and brightly turned on in front of my path allowing me to see everything before I got there.*

7.11.4 Card Sorting

Card sorting showed a near-random distribution of cards in the Circumplex Model of Affect (Table 7.3).

Places	Negative Feelings	Positive Feelings	Total
Calm/Boring	104 (23.2%)	113 (25.2%)	217(48.4%)
Exciting	136 (30.4%)	95 (21.2%)	231(51.6%)
Total	**240 (53.6%)**	**208(46.4%)**	**448**

Table 7.3: Counts of places with respect to the feelings and level of excitement.

However, subjects consistently preferred to liven up what they considered to be the least exciting places with the most interactive lighting modes such as *1 Light Ahead* or *Random*. Subjects preferred the *2 Lights Ahead* mode and the *Always On* mode in places they considered to be dangerous, favoring outlook over entertainment (Figure 7.11).

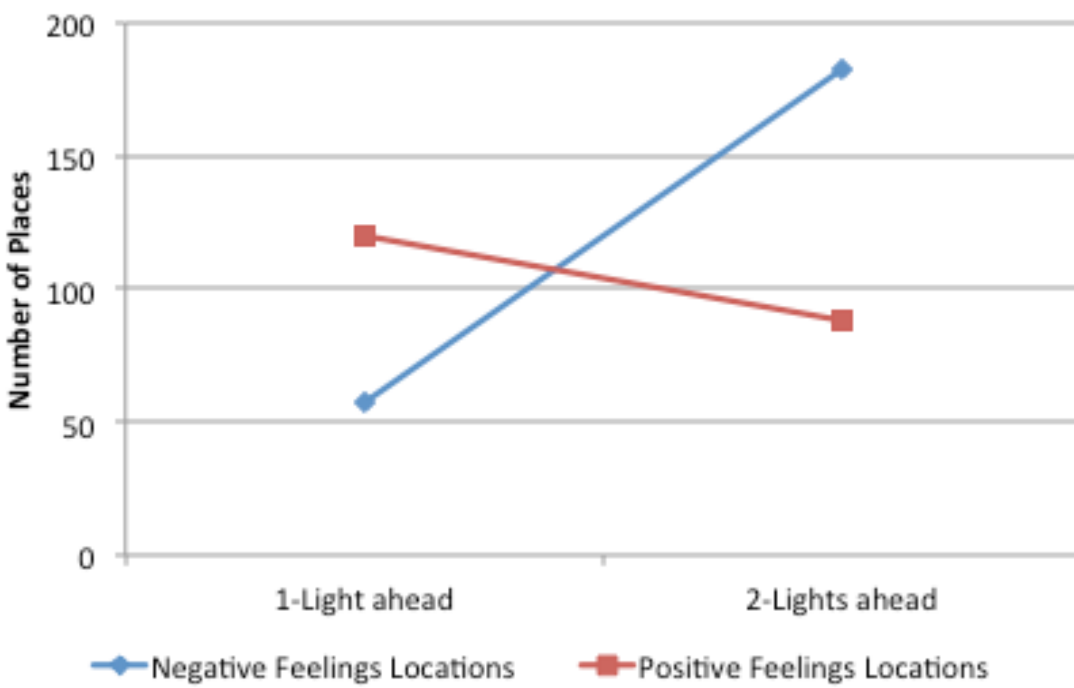


Figure 7.11: Interaction between Valence (feelings) about a place and the number of Look Ahead Lights

The *Ripple* mode was the preferred choice for sites where subjects wished to be warned about obstacles, such as construction sites.

> *If it was an open space, then it didn't feel right to have a base light. Both because there is often already a based light, or because it would pollute the night. If it was a normal space, then no ripple effect. Only if it was a space with some kind of danger, then ripple effect seemed reasonable. If it was somewhere where people moved faster then 2 lights ahead, where they move slower, or rarely, one.*

7.12 **IMPLICATIONS FOR DESIGN**

We have shown that interactive lights modify people's feelings and that the following parameters inform lighting designs that use less energy to "buy" more happiness.

Timing and Ramp Matters

Timing and Ramp affect the way people perceive lights. The *Before* and *Always On* modes are versatile. The *During* mode is useful in places where the user wants to be visible to others, while *Random* and *Ripple* produce some positive emotions only when users feel safe already. Applied in unsafe settings, these same effects produce negative emotions.

Malfunctioning sensor unintended *After* modes. Intended produce negative emotions. A *Fast* transition subjects the sense of more brightness, and meets high visibility expectations better than a *Slow* transition. A *Slow* transition can be used to elicit calm only when no high levels of illumination are expected and when a place is perceived to be safe and carries a positive emotional connotation.

Adaptive Interaction

We found that the level of interactivity must be matched to the level of perceived safety. External factors such as time of the day, neighborhood walkability, as well as internal factors, such as personality, past traumatic life, perceived femininity or masculinity [49] and stress levels modify the perceived safety and therefore the perceived emotion towards a place. These factors modify the subjects' interpretation of a lighting design independently of the designers' intentions. More interactivity introduces broader ranges of affect. (Figure 7.12)

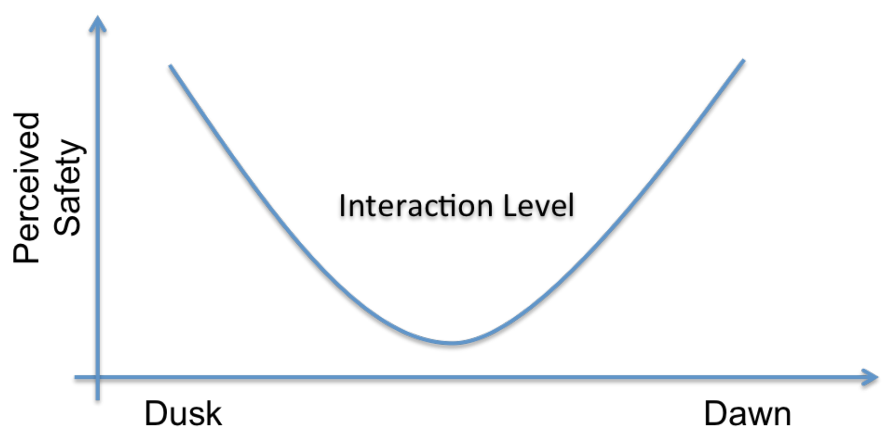


Figure 7.12: Inverse U curve about a place and (a) Timing and (b) Ramp factors.

Energy optimization

When energy optimization is a design objective, interactive lights that trigger positive emotions, such as *Before + Slow* can be used in safe places. On mode; *Base Light* mode with *Before Hard*, *2-lights* ahead mode is a valid alternative to *Always-On* mode even in unsafe conditions. If the perceived safety of a place changes in diurnal or seasonal cycles, the lighting mode can be adapted as needed.

In the context of neighborhood development, interactive lighting systems can be reprogrammed from *Always On* mode to more interactive modes in step with changes in perceived safety, and move from energy-intense to interactive energy-efficient regimes over time. (Table 7.4)

	Positive places (High safety perception)	Negative places (Low safety perception)
Lower Cost of Energy (Fossil Fuel)	**Optimize for Comfort:** - Use a *Soft* ramp with *n-Lights* look ahead as the basic interaction.. and combine with novel interactive modes for long-term engagement. - Reduce illumination to a *Base Light* during low traffic hours; there is no need to use *Always-On*.	**Optimize for Adoption:** - Meet perceived safety levels with an initial *Base Light* or *Always On* roll-out with few *Hard* interactive pilots. - As perceived safety improves, increase the amount and types of interactivity.
Higher Cost of Energy (Renewable)	**Optimize for Energy Consumption:** - Use *Before+Soft* interaction patterns as a default and other *Soft* patterns for entertainment. - Reduce illumination to *Base Light* during low traffic hours and to *None* during peak demand hours.	**Optimize for Low Cost Safety:** - Use *Before+Hard* interactive patterns with *n-Lights* look ahead horizon during dusk or dawn, and *Always On* illumination during darker hours. - Introduce *Soft* interactive patterns as emotions and safety perception improve over the long run.

Table 7.4: Design implications for optimizing affect versus energy consumption in urban interaction illumination systems.

Field deployment

A field roll-out will serve three fundamental purposes: determine the ecological validity of the tool, determine the energetic costs associated with the system, and prove the long term engagement of users. The first two points should be obtained with the actual roll-out, the latter depends on a long term observation and in-situ experimentation. In the Chapter 5 we described how affective interventions suffer from novelty effects despite their efficacy [113]. It is therefore important to commit to not a simple roll-out, but to engage an ecosystem of partners that could help maintain the level of engagement in the system.

Content & API

A key challenge to maintain engagement is to focus on content. We believe content should accommodate three fundamental stakeholders: multimedia and gaming designers, city officials and the actual citizens. We hope to make the system available through a simple mobile interface, for the use of residents as well as through a web API for advanced users.

Multiple users challenge

In the presence of multiple pedestrians, the system has to adapt its interactions according to the amount of traffic. We plan to investigate the challenges and opportunities that we can develop with this system for different traffic scenarios. *Low traffic*, i.e. all pedestrians are sensed individually, can already be attended with some of the same types of light patterns described in this chapter. *Medium traffic*, which contemplates groups of people that interfere with each other should be considered as a special where the interactions could be adapted to treat some groups as individuals and trigger the same interactions at this new level of granularity, or consider novel placements of sensors to improve detection granularity. Finally, for *High traffic*, we need to investigate not only new sensor and actuation configurations, but also external data such as train schedules, traffic light status, ambient devices, social media, wearable computers, and personal data interfaces. Certainly novel sensors and actuation devices could come into play for large groups of people.

Wearable and mobile devices

As already discussed in some of the work presented at the interactive lighting workshops at CHI, NordiCHI, DIS, wearable and personal mobile devices could be a great source of new streams of data that could be used to further improve our research. Affective design could greatly benefit from actual psychophysiological sensors that could generate immediate feedback to the lighting system. For example, a simple detection of emotion through EDA [133], HRV [106], limb movement [135] or body movement [96] could easily help close the loop in terms of emotion management. Furthermore, synchronicity between the system and portable devices could help enhance trigger multimodal interactions with sound and haptics [114].

Design Coherence

The coherence between the affective outcome and the energetic efficiency should be considered by designers. A high level of coherence should be desirable. Coherence should be designed into any multi-modal interaction. If wearables or mobile devices are part of the experience, their expresivity should be coherent with the lights. As observed in chapter 6, poor mode coherence is not neutral. It could generate disgust or negative emotions.

Coverage and Infrastructure

A key issue for IoT interactive systems is coverage. In order to provide the user with a smooth and continuous coverage lights need to be close to each other. Traditional lighting poles are at about 30m, which would render the installation of a smooth path more difficult. Certainly more lights or setting the lights higher would be a possibility. Other options would be to locate the installations sharing public and private infrastructure (i.e. walls). In this case the the anchoring mechanisms should be easy to install and remove.

Vandalism as a design parameter

Sensors and actuators should be replaceable. Design should be fully modular, i.e. if one node is vandalized, the rest should continue to operate. Sensors could have shapes that prevent people from stealing them. Any resemblance of a camera would be a higher temptation. A neutral less appealing design, with a contour, or some texture would be interesting. Drawing inspiration from nature, the shape of the sensor could resemble a wasp nest, or some sharp edges, which are less appealing to humans. Lights should be cheap and highly replaceable. Furthermore, each node should be independent, i.e. tandem or serial installations should be avoided.

7.13 **CONCLUSION**

In this chapter we have shown important and significant relationships between emotions and interactions with urban lights. We identify which design parameters engender optimal positive responses from users. We also showcase the important potential energy consumption and light pollution reduction that can be achieved through precision in interactive designs. Finally, we propose that beyond the well known relationship between illumination and safety, we describe the hierarchical relationship between emotions, more specifically valence, and interaction. We hope that the interactive design parameters such as timing, ramp, look ahead horizon and minimal lighting levels could serve as guidance to future designers who want not only to optimize illumination systems, but increase the overall mood and affective state of urban residents.

8 Conclusion

In this work, we have presented a body of knowledge that describes the importance of doing research in interventions. The amount of data available and the pervasive nature of sensors and actuators drive new interactions. By pushing forward the understanding of bot efficacy as well as improved adoption and attrition reduction, we can describe elements of intervention design such as novelty management, adoption by users (who refuse to be intervened), authoring, ambient interventions, insight discovery and conceptualization. We propose that the use of computation techniques and human-computer interaction goes beyond the traditional aim of automation of psychology, health, and other social sciences. There is a true need to think how to design, computational and affordances are used to create novel interventions that could either support or challenge traditional interventions.

8.5 SUMMARY

8.5.1 Part 1 – Conceptualization and Sensing

In part 1 we laid the ground for enabling intervention design. We described novel ways to discover insights using semantic searching. We leverage a large anonymous database with a broad array of life-related comments to discover new insights. We describe novel approaches to generate concepts for the design of interventions. We base our approach in mixing behavior change and psychological theory with design thinking methods. Finally, we describe a new way of sensing we call "sensorless sensing" based on taking advantage of existing data streams. We describe the opportunity of using regular computer mice movement. We also explore the use of social media as a sensor for affective markers. We propose to use the semantic search insight system as a way to label life events or emotional content.

8.5.2 Part 2 – Opportunistic Interventions

In part 2 we focus on different intervention challenges based on opportunistic approaches. We show that mobile and wearable interventions are as powerful as traditional ones, such as visualization. Then we describe a novel use of web apps to deliver micro-interventions based on contextual and personal information. We describe how to "harvest" interventions from the web. We could potentially generate a large database of micro-interventions based on popular media. We outline the challenges of creating coherent multimodal affective interactions. We describe the use of ambient technologies, such as LED lights. We show that we can create positive affect reactions with urban interactive lights. We describe the inverse relationship between the feelings about places and interactivity. Finally, we conclude that optimal affective interactive systems can also be more energy efficient.

8.6 INTERVENTION RESEARCH DRIVERS

To potentialize intervention research we should be aware that:

a. *Engaging interventions are a Human-Centered Design challenge*. Proper interventions are NOT an organic evolution of sensing research. Conceptualization demands a knowledge of the history and many times a deconstruction of reality. Furthermore, valid concepts should focus on design but also efficacy.

b. *Multi-everything*. We need the multiplicity of sensors and actuators, a multi-disciplinary approach, and suites of interventions. There is not a "one-size-fits-all" solution. The most effective and efficient interventions should be able to multi-faceted in every axis.

c. *Adoption is not trivial.* Good design is NOT cheap, NOR easy. It requires much effort, talent, and investment. Making a successful intervention out of a sensing stream or from a sound theoretical base is not trivial. We should look for opportunities to "harvest" good design or establish fluid authoring systems. Furthermore, even if we have the right intervention, we may not appeal the right audience. In some cases, we need to design ease-in phases. In other we may need to build disguised solutions to bring people on board.

d. *Attrition is higher in technology*. Simplicity to access, novelty, and word-of-mouth marketing drive many users to try new interventions. We should acknowledge that a substantial percentage will drop unavoidably due to unmatched expectations. The remaining ones are part of the population that we should maintain and care. Furthermore, some attrition is "desirable"; these are the "learners"; people that learn from our system and do NOT need to keep using it. Metaphorically speaking, we decide if we are designing "crutches" or "wheelchairs" for well-being.

8.7 VISION

Well-being technology requires a change in the approach to the design of novel technologies. We need to focus on hard but fundamental challenges to make this technology happen. We list here a few:

a. *Individual empowerment*. We need to learn by contrasts and throughout time. Different timelines and different individuals can support the creation of novel suites of interventions. Furthermore, long-term analysis of data could help us mine "archetypes". If we find the "other me" I could learn from her/him how she/he recovered from a hardship. We need to use data-driven concepts and prototypes, but with a precision for a N=1 population. Only creative concepts that are personalizable can bridge the chasm of adoption.

b. *Cultivate stable "Equilibrium".* We need to understand what are our values (i.e. our invariants). We can use our values and strengths to keep us stable, resilient.

A “value engineering” approach could lead to longer term behavior change. Blending values could be a good way to overcome resistance to change.

c. *Ecosystems*. We need on-demand, passive, ambient devices that are always ready to give us good advice. We should design personal and ambient systems that are “coherent” with our emotions.

d. *Interpersonal relationships*. Pervasive Well-being Technology should solve social problems at the personal level and vice-versa. Only our designs work well for groups we could say we have reached maturity.

References

1. Alagarai Sampath, H., Indurkhya, B., Lee, E., & Bae, Y. (2015, April). Towards Multimodal Affective Feedback: Interaction between Visual and Haptic Modalities. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2043-2052). ACM.
2. Aliakseyeu, D., Meerbeek, B., Mason, J., Lucero, A., Ozcelebi, T., & Pihlajaniemi, H. (2014, October). Beyond the switch: explicit and implicit interaction with light. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (pp. 785- 788). ACM.
3. Aliakseyeu, D., Meerbeek, B., Mason, J., van Essen, H., Offermans, S., Wiethoff, A., & Lucero, A. (2012, June). Designing interactive lighting. InProceedings of the Designing Interactive Systems Conf.(pp. 801-802). ACM.
4. Aliakseyeu, D., van Essen, H., Lucero, A., Mason, J., Meerbeek, B., den Ouden, E., & Wiethoff, A. (2013, April). Interactive city lighting. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 3191-3194). ACM.
5. American Psychological Association, "Stress in America 2010", Retrieved December 13, 2010 from: http://www.apa.org/news/press/releases/stress/
6. Atlas.ti, "Atlas.ti – Qualitative Data Analysis", Retrieved December 17, 2015 from: http://atlasti.com/
7. Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. Machine learning, 47(2-3),235-256.
8. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, Vol 84(2), 191-215.
9. Banerjee, S.P., and Woodard, D.L. (2012). Biometric Authentication and Identification Using Keystroke Dynamics: A Survey. Journal of Pattern Recognition Research, 7 (1), (2012), 116-139.
10. Baranowski, T., Buday, R., Thompson, D., Baranowski, J. (2008). Playing for Real: Video Games and Stories for Health-Related Behavior Change. American Journal of Preventive Medicine, Vol 34(1), 74-82

11. Beating the Blues, "Beating the Blues – cognitive behavioral therapy", Retrieved December 17, 2015 from: http://www.beatingtheblues.co.uk/

12. Beck, A. (1970, May). Cognitive Therapy: Nature and Relation to Behavior Therapy; Behavior Therapy, Volume 1, Issue2, Pages 184-200

13. Blei, D. (2012, April). Probabilistic topic models. *Communications ACM* 55, 4, 77–84.

14. Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?, American psychologist, 59(1), 20.

15. Bradley M., and Lang, P. (2007). "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual," Technical Report B-3, University of Florida.

16. Breiman, L. and Cutler, A. Random forests-classification description. Department of Statistics, Berkeley

17. Brindley, D. N., & Rolland, Y. (1989). Possible connections between stress, diabetes, obesity, hypertension and altered lipoprotein metabolism that may result in atherosclerosis. *Hypertension*, *23*, 33-351.

18. Brown L. and Williamson, J. (2007). "Shake2Talk: Multimodal Messaging for Interpersonal Communication," *Haptic and Audio Interaction Design*, vol. 4813, pp. 44-55.

19. Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. arXiv preprint arXiv:1204.5721.

20. Burchfield, S. (1979). The stress response: a new perspective, American Psychosomatic Society

21. Butler, A. C., Chapman, J. E, Forman, E. M. and Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. Clinical Psychology Review, 26(1):17-31

22. Cahill, L., McGaugh, J. (1995). A Novel Demonstration of Enhanced Memory Associated with Emotional Arousal, Consciousness and Cognition 4, 410-421, © Acad. Press, Inc.

23. Carson, J., & Kuipers, E. (1998). Stress management interventions, Occupational stress: Personal and professional approaches, 157-74.

24. Chang, K., Fisher, D., Canny, J., and Hartmann, B. (2011). How's my mood and stress?: an efficient speech analysis library for unobtrusive monitoring on mobile phones. In Proceedings of the 6th International Conference on Body Area Networks, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 71–77.

25. Christianson, S. (1992, September). Emotional stress and eyewitness memory: A critical review. Psychological Bulletin, Vol 112(2), 284-309

26. Cohen, S., Kessler, R., Underwood L. (1995). Strategies for Measuring Stress in Studies of Psychiatric and Physical Disorders, Oxford University Press.

27. Consolvo, S., Klasnja, P., McDonald, D., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher K. and Landay, L. (2008). Flower or a Robot Army? Encouraging Awareness and Activity with Personal, Mobile, Displays, International Conference on Ubiquitous Computing.

28. Cooper, E. W., Kryssanov, V. V., & Ogawa, H. (2010). Building a framework for communication of emotional state through interaction with haptic devices. In *Haptic and Audio Interaction Design* (pp. 189-196). Springer Berlin Heidelberg

29. Coyle, D., Matthews, D., M., Sharry, J., Nisbet A., and Doherty, G. (2005). Personal Investigator: A Therapeutic 3D Game for Adolescent Psychotherapy, International Journal of Interactive Technology and Smart Education, 2(2):73-88.

30. Coyle, D., McGlade, N., Doherty, G. & O'Reilly, G. (2011) Exploratory Evaluations of a Computer Game Supporting Cognitive Behavioural Therapy for Adolescents. ACM CHI 2011

31. Dietz P. H., Eidelson B. D., Westhues J., and Bathiche S.A. (2009). Practical pressure sensitive computer keyboard. In UIST, 55-58.

32. Djuric, N., Wu, H., Radosavljevic, V., Grbovic, M. and Bhamidipati, N. (2015). Hierarchical neural language models for joint representation of streaming documents and their content. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. 248–255.

33. Doherty, G., Coyle, D., & Sharry, J. (2012, May). Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In Proc. of the 2012 ACM annual conference on Human Factors in Computing Systems

(pp. 1421-1430). ACM.

34. Duric, Z., Gray, W. D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M. J., ... & Wechsler, H. (2002). Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. Proceedings of the IEEE, 90(7), 1272-1289.

35. Eisenstein, J., Horng Chau, D., Kittur, A. and Xing, E. (2012). TopicViz: Interactive topic exploration in document collections. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 2177–2182.

36. Ekman, P. (1993). Facial expression and emotion. *American psychologist*,*48*(4), 384.

37. Epp, C., M. Lippold, and R.L. Mandryk. (2011). Identifying Emotional States using Keystroke Dynamics, in ACM Conference on Human Factors in Computer Systems CHI'2011: Vancouver, BC.

38. Eysenbach, G. (2005). The law of attrition. Journal of medical Internet research, 7(1).

39. Fearfighter, "Fearfighter", Retrieved December 17, 2015 from: http://www.fearfighter.com/

40. Fischer, M. (1966). The KWIC index concept: A retrospective view. *American Documentation* 17, 2, 57–70.

41. Fitbit, "Fitbit Store", Retrieved December 17, 2015 from: http://www.fitbit.com/

42. Fotios, S., and C. Cheal. (2012). "Using obstacle detection to identify appropriate illuminances for lighting in residential roads." Lighting Research and Technology.

43. Fotios, S., Unwin, J., & Farrall, S. (2014). Road lighting and pedestrian reassurance after dark: A review. Lighting Research and Technology, 1477153514524587.

44. Freesound.org, "Freesound.org", Retrieved December 17, 2015 from: https://www.freesound.org.

45. Gabrilovich, E. and Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*. San Francisco, CA, USA, 1606–1611.

46. Gaffary, Y., Eyharabide, V., Tsalamlal, M. Y., Martin, J. C., & Ammi, M. (2013). Comparison of Statistical Methods for the Analysis of Affective Haptic Expressions. In *Haptic and Audio Interaction Design* (pp. 69-78). Springer Berlin Heidelberg.

47. Gao Y., Bianchi-Berthouze N., and Meng H. (2012). What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay? In ACM Trans. CHI. 19 (4).

48. Green, S., and Bavelier, D., (2003). Action video games modifies visual selective attention. Nature, 423, 534-537

49. Haans, A., & de Kort, Y. A. (2012). Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, prospect, concealment, and escape. Journal of Environmental Psyc., 32(4), 342-352.

50. Hanyu, K. (1997). Visual properties and affective appraisals in residential areas after dark. Journal of Environmental Psychology, 17(4), 301-315.

51. Harrison, C., Horstman, J., Hsieh, G., & Hudson, S. (2012, May). Unlocking the expressivity of point lights. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1683-1692). ACM.

52. Healey, J., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. Intelligent Transportation Systems, IEEE Transactions on, 6(2), 156-166.

53. Heartmath, "Heartmath store", Retrieved December 17, 2015 from: http://www.heartmathstore.com/

54. Heinrichs, M., Baumgartner, T., Kirschbaum C. and Ulrike E. (2003, April). Social Support and Oxytocin Interact to Suppress Cortisol and Subjective Responses to Psychosocial Stress; 2003 Society of Biological Psychiatry, Zurich, Switzerland

55. Hernandez, J., Paredes, P., Roseway, A., & Czerwinski, M. (2014, April). Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 51-60). ACM.

56. Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion*, *6*(3), 528.

57. Hofer, A., Siedentopf, C. M., Ischebeck, A., Rettenbacher, M. A., Verius, M., Felber, S., & Fleischhacker, W. W. (2006). Gender differences in regional cerebral activity during the

perception of emotion: a functional MRI study.*Neuroimage*, *32*(2), 854-862.

58. Hong, J., Ramos, J., and Dey, A. (2012). Understanding physiological responses to stressors during physical activity. In Proceedings of Ubicomp.

59. Hsu, S., Lee, F., and Wu, M. (2005). Designing Action games for appealing to buyers. Cyberpshychology Behavior, 8:585-91

60. Hudlicka, E. and McNeese, M. (2002). Assessment of User Affective and Belief States for Interface Adaptation: Application to an Air Force Pilot Task, User Modelingand User-Adapted Interaction.

61. Huisman, G., Darriba Frederiks, A., Van Dijk, B., Hevlen, D., & Krose, B. (2013, April). The TaSST: Tactile sleeve for social touch. In World Haptics Conference (WHC), 2013 (pp. 211-216). IEEE.

62. Ishii, H. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms, ACM - CHI '97

63. Israr, A., & Poupyrev, I. (2011, May). Tactile brush: drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2019-2028). ACM.

64. Ju, W., & Takayama, L. (2009). Approachability: How people interpret automatic door movement as gesture. International Journal of Design, 3(2), 1-10.

65. Ju, W., Lee, B. A., & Klemmer, S. R. (2008, November). Range: exploring implicit interaction through electronic whiteboard design. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (pp. 17-26). ACM.

66. Kalia, M. (2002). Assessing the economic impact of stress - The modern day hidden epidemic. Metabolism, 51(6), 49-53.

67. Kato, P. (2010). Video Games in Health Care: Closing the Gap, Review of General Psychology, 2010, Vol 14., No. 2, 113-121

68. Kenter, T. and de Rijke, M. (2015). Short text similarity with word embeddings. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management.*, Vol. 15. 115.

69. Khan I. A., Brinkman W., Fine N., and Hierons R. M. (2008). Measuring personality from keyboard and mouse use. In Cognitive ergonomics: the ergonomics of cool interaction, 38.

70. Khanna P., and Sasikumar M. (2010). Recognizing Emotions from Keyboard Stroke Pattern. In International Journal of Computer Applications, 11(9), 2010, 1-5.

71. Kiropoulous, L. A., Klein, B., Austin, D. W., Gilson, K., Pier, C., Mitchell, J. and Ciechomski, L. (2008). Is internet based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? Journal of Anxiety Disorders, 22: 1273-1284.

72. Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Skip-thought vectors. In Advances in Neural Information Processing Systems (pp. 3276-3284).

73. Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. Psychiatry Ann., 32(9), 1-7.

74. Kurihara, K. (2007). ed. *Face Recognition Book*. Pro Literatur Verlag.

75. Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. Springer Publishing Company.

76. Living Life to the Full, “Living Life to the Full”, Retrieved December 17, 2015 from: http://www.llttf.com/

77. Lu, H., Rabbi, M., Chittaranjan, G., Frauendorfer, D., Mast, M., Campbell, A., Gatica-Perez, D., and Choudhury, T. (2012). Stresssense: Detecting stress in unconstrained acoustic environments usingsmartphones. 14th Intl Conf. Ubiquitous Computing

78. Luhn. H.P. (1960). Key word-in-context index for technical literature (KWIC index). *American Documentation* 11, 4 (1960), 288–295.

79. Lumosity, “Lumosity”, Retrieved December 17, 2015 from: http://www.lumosity.com

80. MacKenzie, S. (1992). Fitts' law as a research and design tool in human-computer interaction. In Hum.-Comput. Interact. 7, 1, (1992), 91-139.

81. MacLean, D., Roseway, A., & Czerwinski, M. (2013, May). MoodWings: a wearable biofeedback device for real-time stress intervention. In Proc. of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (p.

66). ACM.

82. Maehr, W., (2005). eMotion: Estimation of the User's Emotional State by Mouse Motions, in Information and Communication Engineering 2005, Fachhochschule Vorarlberg: Dornbirn, Austria.

83. Malone, T.W. (1980). What Makes Things Fun to Learn?: Heuristics for Designing Instructional Computer Games. In Proc. SIGSMALL, ACM Press, NY, USA, 162-169

84. Marks, I. M., Kavanaugh, K. and Gega, L. (2007). Hands-On Help: Computer-Aided Psychotherapy. Psychology Press

85. Martínez Garcia, E., España-Bonet, D. and Màrquez., L. (2015). Document-level machine translation with word vector models. In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*. Antalya, Turkey, 59–66.

86. MaxQDA, "MAXQDA: The Art of Text analysis", Retrieved September 17, 2015 from: http://www.maxqda.com/index.php/maxqda

87. Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. Journal of marketing research, 45(6), 633-.

88. McEwen, B. S., & Stellar, E. (1993). Stress and the individual: mechanisms leading to disease. Archives of internal medicine, 153(18), 2093.

89. Mihalcea, R. and Csomai, A. (2007). Wikify!: Linking documents to encyclopedic knowledge. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management (CIKM '07)*. New York, NY, USA, 233–242.

90. Mihalcea, R., Corley, C. and Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1 (AAAI'06)*. 775–780.

91. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

92. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

93. Mikolov., T. word2vec: Tool for computing continuous distributed representations of words, Retrieved September 23, 2015 from https://code.google.com/p/word2vec/.

94. Miller, G. (2006). The unseen: Mental illness's global toll. Science, 313:458–461.

95. Mohr, D. C., Vella, L., Hart, S., Heckman, T. and Simon, G. (2008). The effect of telephone-administered psychotherapy on symptoms of depression and attrition: A meta-analysis. Clinical Psychology: Science and Practice, 15(3):243–253.

96. Monrose, F., and Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems, 13, 351-359.

96. Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. Journal of Nonverbal Behavior, 11(1), 33-42.

98. Moodgym, "Moodgym", Retrieved December 17, 2015 from: http://moodgym.anu.edu.au/welcome

99. Moraveji, N., Olson, B., Nguyen, T., Saadat, M., Khalighi, Y., Pea, R., & Heer, J. (2011, October). Peripheral paced respiration: influencing user physiology during information work. In Proc. of the 24th annual ACM symposium on User interface software and technology. ACM.

100. Muñoz, R. F., Lenert, L. L., Delucchi, K., Stoddard, J., Perez, J. E., Penilla, C., & Pérez-Stable, E. J. (2006). Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. Nicotine & Tobacco Research, 8(1), 77-87.

101. Munoz, R. (2000). Group Cognitive Behavioral Therapy of Major Depression, Latino Mental Health Research Program, UCSF

102. Muñoz, R., Mendelson, T. (2005). Toward Evidence-Based Internet Interventions for Diverse Populations: The San Francisco General Hospital Prevention and Treatment Manuals, Journal of Consulting and Clinical Psy. Vol. 73, No. 5, 790-799

103. Muralidharan A., and Hearst, M. (2013). Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing* 28, 2 (2013), 283–295.

104. Muralidharan, A., Hearst, M., and Fan, C. (2013). WordSeer: A knowledge synthesis environment for textual data. In *Proceedings of the 22Nd ACM International Conference on*

Conference on Information & Knowledge Management (CIKM '13). ACM, New York, NY, USA, 2533–2536.

105. Murray, C. and Lopez, A., (1996). editors. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Harvard University Press.

106. Naima, R., & Canny, J. (2009, June). The berkeley tricorder: Ambulatory health monitoring. In Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on (pp. 53-58). IEEE.

107. Nike plus, "Nike Plus", Retrieved December 17, 2015 from: http://nikeplus.nike.com/plus/

108. nVivo, "QSR: nVivo", Retrieved December 17, 2015 from: http://www.qsrinternational.com/product

109. Offermans, S. A. M., van Essen, H. A., & Eggen, J. H. (2014). User interaction with everyday lighting systems. Personal and Ubiquitous Comp., 18(8), 2035-2055.

110. Okuda, S., Ishii, J., & Fukagawa, K. (2007). Research on the Lighting Environment in the Street at Night – Part 1. Resident's attitude on the safety and security in the street at night. Proceedings of 26th Session of the CIE (Volume 2).

111. Olsen Jr, D. R. (2007, October). Evaluating user interface systems research. In Proc. of the 20th annual ACM symposium on User interface software and technology (pp. 251-258). ACM.

112. Paredes, P. and Chan, M. (2011) CalmMeNow: exploratory research and design of stress mitigating mobile interventions. SIGCHI 2011 Extended Abstracts on Human Factors in Computing Systems. ACM

113. Paredes, P., Gilad-Bachrach, R., Czerwinski, M., Roseway, A., Rowan, K., & Hernandez, J. (2014, May). PopTherapy: Coping with stress through pop-culture. In Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (pp. 109-117). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

114. Paredes, P., Ko, R., Aghaseyedjavadi, A., Babler, L., Chuang, J., Canny, J., Synestouch, Haptic + Audio (2015, September). Affective Design for Wearable Devices, Proceedings of the 6th Affective Computing and Intelligent Interaction, ACII2015. IEEE

115. Persons, J. (1989). Cognitive Behavioral Therapy in Practice: A Case Formulation Approach

116. Picard, R. (2003). Affective computing: challenges. International Journal of Human-Computer Studies 59, 1, 55–64.

117. Pirhonen, A., & Tuuri, K. (2011). Calm Down–Exploiting Sensorimotor Entrainment in Breathing Regulation Application. In *Haptic and Audio Interaction Design* (pp. 61-70). Springer Berlin Heidelberg.

118. Poh, M.-Z., Mcduff, D.J. and Picard, R. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express. 18(10).

119. Proceedings of the HAID - Haptic and Audio Interaction Design International Workshops, 2006 - 2013

120. PTSD Coach, "PTSD Coach", Retrieved December 17, 2013 from: http://www.ptsd.va.gov/public/pages/PTSDCoach.asp

121. Rattenbury, T. and Canny, J. (2007). Task Support using Activity-Based Context. in ACM Conference on Human Factors in Information Systems, CHI'07

122. Relax to win, "Relax to win", Retrieved December 17, 2013 from: http://gizmodo.com/102685/philips-relax-to- win-sensor

123. Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: a meta-analysis. Journal of occupational health psychology, 13(1), 69.

124. Richter, H., Blaha, B., Wiethoff, A., Baur, D., & Butz, A. (2011, September). Tactile feedback without a big fuss: simple actuators for high-resolution phantom sensations. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 85-88). ACM.

125. Rosenstiel, A. K., & Keefe, F. J. (1983). The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain, 17(1), 33-44.

126. Routasalo, P., & Isola, A. (1996). The right to touch and be touched. Nursing Ethics, 3(2), 165-176.

127. Russell, J. A. (1980). A circumplex model of affect. Journal of

Personality and Social Psychology, 39, 1161-1178

128. Ryan, R., Rigby, C., Przbylski, A. (2006). The motivational pull of videogames: a self determination theory approach. Motivation and Emotion. 347-63.

129. Scanlon, M., Drescher, D., Sarkar, K. (2007). Improvement of Visual Attention and Working Memory through a Web-based Cognitive Training Program, Lumos Labs.

130. Scheier, M. F., Weintraub, J. K., & Carver, C. S. (1986). Coping with stress: divergent strategies of optimists and pessimists. Journal of personality and social psychology, 51(6), 1257.

131. Schueller, S. M. (2010). Preferences for positive psychology exercises. The Journal of Positive Psychology, 5(3), 192-203.

132. Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: an introduction. American psychologist, 55(1), 5.

133. Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G., & Picard, R. W. (2005). The handwave bluetooth skin conductance sensor. In Affective computing and intelligent interaction (pp. 699-706). Springer Berlin Heidelberg.

134. Stürmer, S., Snyder, M., & Omoto, A. M. (2005). Prosocial emotions and helping: the moderating role of group membership. *Journal of personality and social psychology*, *88*(3), 532.

135. Sun, D., Paredes, P., & Canny, J. (2014, April). MouStress: detecting stress from mouse motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 61-70). ACM.

136. Superbetter, “Superbetter.us”, Retrieved December 17, 2013 from: https://www.superbetter.com/

137. Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, *17*, 354-381.

138. Tian, Y., Kanate, T. and Cohn, J. (2001). Recognizing action units for facial expression analysis, IEEE Trans. Pattern Analysis Machin Intelligence, 23 (2), pp. 97-115

139. Tikka, V., & Laitinen, P. (2006). Designing haptic feedback for touch display: Experimental study of perceived intensity and integration of haptic and audio. In *Haptic and Audio Interaction Design* (pp. 36-44). Springer Berlin Heidelberg.

140. Top Web Apps, "Alexa", Retrieved December 17, 2015 from: http://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Web_Applications

141. Turney., P. (2001). Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the 12th European Conference on Machine Learning (ECML '01)*. Springer-Verlag, London, UK, UK, 491–502.

142. tvlight.com, "TVlight: Empowering Intelligence", Retrieved December 17, 2015 from: http://www.tvilight.com/

143. Unwin, J., & Fotios, S. (2011). Does lighting contribute to the reassurance of pedestrians at night-time in residential roads?. *Ingineria Iluminatului*, *2*(13), 29-44.

144. Vaucelle, C., Bonanni, L. and Ishii, L. (2009, April). Design of Haptic Interfaces for Therapy; CHI'09, Boston, MA, US

145. Vickers, P. (2013). Ways of listening and modes of being: Electroacoustic auditory display. arXiv preprint arXiv:1311.5880.

146. Villar, N., Izadi, S., Rosenfeld, D., Benko, H., Helmes, J., Westhues, J., Hodges, S., Butler, A., Ofek, E., Cao, X., and Chen, B. (2009). Mouse 2.0: Multi-touch Meets the Mouse. In User Inter. Soft. and Tech., 33-42.

147. Vizer L. M., Zhou L., and Sears A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. Int. J. Hum.-Comput. Stud. 67, 10, 870-886.

148. Watson, D., Clark, L. and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales." Journal of personality and social psychology 54.6

149. Wattenberg M., and Viegas, F.B. (2008). The Word Tree, an interactive visual concordance. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1221–1228.

150. Weierich, M. R., Wright, C. I., Negreira, A., Dickerson, B. C., & Barrett, L. F. (2010). Novelty as a dimension in the affective brain. *Neuroimage*, *49*(3), 2871-2878.

151. White, M. (1995). Re-authoring Lives: Interview and Essays, Adelaide, South Australia: Dulwich Centre Publications.

152. Windows Phone, "Windows Phone Top Rated Apps", Retrieved December 17, 2013 from: http://www.windowsphone.com/en-us/store/top-rated-apps

153. Windows Phone, “Windows Phone Top Rated Games”, Retrieved December 17, 2013 from: http://www.windowsphone.com/en-us/store/top-rated-games

154. Wood, J. V., Saltzberg, J. A., Neale, J. M., Stone, A. A., & Rachmiel, T. B. (1990). Self-focused attention, coping responses, and distressed mood in everyday life. *Journal of personality and social psychology*, *58*(6), 1027.

155. Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*,*8*(4), 494.