Using Interactive Genetic Algorithms to Support Aesthetic Ergonomic Design

by

Dan Nathan-Roberts

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Industrial and Operations Engineering) in The University of Michigan 2012

Doctoral Committee:

Professor Yili Liu, Chair
Professor Roman Kapuscinski
Associate Professor Bernard J. Martin
Assistant Research Scientist Jarod C. Kelly


Motorisches Zentrum
Kopf Beine Arme
Wille
N
THE UNIVERSITY OF MICHIGAN
ARTES SCIENTIA VERITAS
1817

DEDICATION

To my mother and father, my partner Julia, my uncle Marty, and my friends, without whose support this would not have been possible.

Meliora

ACKNOWLEDGEMENTS

It is with deep gratitude that I would like to acknowledge the contributions, support, wisdom, forgiveness, and guidance that I have received from so many people that enabled me to get here.

I would like to acknowledge the contributions of my father, Peter for his support, and tireless encouragement, my mother, Miriam for her wise counsel, and the design of the frontispiece of this dissertation, my partner, Julia for her understanding and encouragement, my uncle Martin J. Nathan for his support throughout my life, and the rest of my family who have served as role models of informed citizenry.

My Advisor, Dr. Yili Liu has provided me honest, often poetic advice, for which I am very thankful. In addition to his advice, I appreciate his work and support throughout my dissertation. I would also like to acknowledge my dissertation committee, Dr. Jarod Kelly, for his assistance with Interactive Genetic Algorithms, Dr. Bernard Martin, for his training on biomechanics and feedback on my work, and Dr. Roman Kapuscinski for his support and advice throughout my graduate career.

There have been a number of teachers who have had an impact on my pursuits that went far beyond their courses, including Mr. Davis, Mr. Olivier, Mr. Davis, Mr. Naudé, Mrs. Hack, and others at Saint Mary's College High School. University of Rochester professors Dr. Quesnel, Dr. Burns, Dr. Aslin, Barry Gilbert, Dr. Eastman, Curt Smith, and Sensei Hughes among others left a lasting impact as well. At the University of Michigan, Dr. Kantowitz, Dr. Sarter, Dr. Degraff, Dr. Green, Dr. Seiford, and a number of other professors across the university have had a hand in shaping the researcher I am becoming.

Outside of school, I have been fortunate to receive mentoring and training from a number of exceptional and dedicated people including Dr. David Rempel, who helped me discover ergonomics and develop as a researcher, Dr. Sandy Weininger, who inculcated in me a passion for healthcare, Jim Kellso, who helped me develop an individual vision, and Thomas Mann, who trained me as an engineer. Krishna Asundi, Ed Young, and Kristin Amlie through the University of California Ergonomics Program, and Candice Bauer, and Mohammad Mahinfalah through ASME have provided significant mentoring and support as well, for which I am very grateful.

I am confident that it would not have been as easy to graduate from the University of Rochester, let alone achieve Ph.D. candidacy at the University of Michigan without the kindness, generosity, support, and friendship of the outstanding staff that I have been fortunate enough to interact with, including Carla, Carm, Elayne, Jill, Lisa, and Rosemary at the University of Rochester, Betsy at the University of California Ergonomics Program, Cecilia, Shannon, and Marian in ASME, Amy, Chris, Chuck, Diana, Eyvind, Liz, Matt, Maureen, Michaele, Mint, Rachel, Rick, Rod, Sarah, Tershia, Tina, and Wanda here at the University of Michigan.

I am eternally grateful for friends who encouraged me to be an engineer and accompanied me on similar paths including Mookie, Chris, J.J., Erin, Ari, Dave, The Family, Ramin, Eduardo, and many other classmates and friends. The friendship and support from past Ph.D. students while at the University of Michigan including Katrina Appell, Kristi Schmidt Bauerly, Michael Bauerly, Thomas Ferris, and Michael Lau has meant a lot to me. While in graduate school, I have had the support of a number of close friends including Andrew, Arleigh, Gökçe, Julie, Neal, Bob Blackburn, Bob Hospadaruk, Seymour Greenstone, Shanna, and many other close friends for whom I wish there was space here. Additionally, this research was made possible with the help of a number of undergraduate students including Alex Beeker, Christian Little, Christine An So, Jon D. Hays, Liang Hao Zhao, and Michael Young.

TABLE OF CONTENTS

DEDICATION ii
ACKNOWLEDGEMENTS iii
LIST OF FIGURES viii
LIST OF TABLES x
LIST OF EQUATIONS xii
ABSTRACT xiii
Chapter 1 Introduction to Aesthetic and Ergonomic Design of Mobile Phones and Blood Glucose Meters 1
- Aesthetic Ergonomics 1
- Dual Process Research Methodology 2
- Modeling Aesthetics 2
- Physical Ergonomics 4
- Modeling Physical Ergonomics 5
- Interactive Genetic Algorithms 6
- Genetic Algorithms 8
- Research Domains 10
  - Mobile Phones 11
  - Blood Glucose Meters 12
- References 15

Chapter 2 Content Analysis, Anthropometric Modeling, and Exploratory IGA study of Aesthetic Ergonomics 18
- Introduction 18
- Study One: Content Analysis of Literature on Mobile Phone Design Features 18
  - Method 18
  - Results 19
  - Discussion 21
- Study Two: Physical Ergonomic Modeling of Two Problems with Mobile Phones 22
  - Method 22
  - Results 24
  - Discussion 25
- Study Three: Exploratory Study of Mobile Phone Aesthetics using an IGA 26
  - Method 26
  - Results 30
  - Discussion 34

References 37

Chapter 3 Testing Aesthetic and Function Design Preference for Touchscreen and Non-Touchscreen Mobile Phones Using Interactive Genetic Algorithms 38
Introduction 38
Goals 38
Chapter Three Experiment One 39
Method 39
Results 43
Discussion 48
Chapter Three, Experiment Two 51
Method 51
Results 53
Discussion 56
General Discussion 57
References 60

Chapter 4 Comparison of Design Preferences for Mobile Phones and Blood Glucose Meters for Experienced and Non-User Groups 61
Introduction 61
Goals 61
Chapter Four, Experiment One 61
Method 61
Results 67
Discussion 69
Chapter Four, Experiment Two 70
Method 70
Results 72
Discussion 77
General Discussion 79
References 82

Chapter 5 Integrating Aesthetic and Usability Factors in the Design of Mobile Phones and Blood Glucose Meters 85
Introduction 85
Goals 85
Chapter Five, Experiment One 85
Method 85
Results 91
Discussion 92
Chapter Five, Experiment Two 94
Method 94
Results 96
Discussion 97

General Discussion 98
References 101

Chapter 6 Summary of Findings 102
Summary of Aesthetics Findings 102
How to Modulate Aesthetic Preference 102
Cross-Domain Ties of Aesthetic Preference Findings 104
Summary of Mobile Device Specific Aesthetic Findings 105
Summary of Physical Ergonomics Findings 106
Genetic Algorithms 106
Summary of Mobile Device Specific Physical Ergonomics Findings 107
Summary of Findings Related to Aesthetic Ergonomics Research Methods 107
Summary of Interactive Genetic Algorithm Specific Findings 107
Future Directions 108
Future Aesthetics Research Directions 108
Future Physical Ergonomics Research 109
Future Tools and Methods 109
Conclusion 110
References 112

LIST OF FIGURES

Figure 1.1. IGA Process diagram, showing how variables make a population of designs and user input as well as the genetic algorithm iteratively evolve that design (Kelly J. C., 2008).. 7
Figure 1.2. Sample of commercially available blood glucose meters. 13
Figure 2.1. Mobile phone design using mean measured values reported in Table 2.3. This phone is sized to have the mean proportions of all of the phones measured in the literature. 21
Figure 2.2. Simple anthropometric model relating phone length to elbow angle. This simplified model makes several assumptions about posture, and serves as an illustrative posture only.. 23
Figure 2.3. Elbow angle increase as a function of phone length. 24
Figure 2.4. Mean of Arial and Times New Roman maximum legible distance for three characters densities on a square screen.. 25
Figure 2.5. Independent variables being tested.. 27
Figure 2.6. Screen capture of the software tool. The four phones with dark backgrounds around the screen and buttons indicate they are currently selected by the user. 29
Figure 2.7. Final selected mobile phone designs, each design being the design selected on the tenth generation by a participation as the most aesthetically pleasing. 31
Figure 2.8. Standard Deviation (SD) of the horizontal and vertical screen dimensions by generation. The entire population's SD and the selected phone’s SD are shown. 32
Figure 2.9. Entire and selected populations’ Standard Deviation (SD) of the horizontal and vertical button spacing as well as the phone corner radius by generation. 32
Figure 2.10. Variables plotted by loadings on components 1 and 2. 34
Figure 3.1. Independent dimension variables in mobile phone design............................ 40
Figure 3.2. Screen capture of IGA interface with four mobile phones selected, shown with a darker phone background.. 42
Figure 3.3. Standard deviation decrease by generation for the variables listed (selected designs and population as a whole). 44
Figure 3.4. Standard deviation decrease by generation for the variables listed (selected designs and population as a whole). 45
Figure 3.5. Graph showing mean values. Connecting lines indicate significantly different in pairwise comparisons at $\alpha = 0.0033$ (corrected with Bonferroni). Error bars indicate standard deviation. 46
Figure 3.6. Graph showing mean values. Connecting lines indicate significantly different in pairwise comparisons at $\alpha = 0.0033$ (corrected with Bonferroni). Error bars indicate standard deviation. 46

Figure 3.7. Graph showing mean values. Connecting lines indicate significantly different mean values in pairwise comparisons at $\alpha = 0.01$ (corrected with Bonferroni). Error bars indicate standard deviation. ... 54
Figure 3.8. Mean design for each prompt separated by the type of phone owned by participant showing touchscreen phone owners preferred taller keypads and statistically significantly taller screens. ... 56
Figure 4.1. Independent dimension variables in mobile phone design. ... 62
Figure 4.2. Independent dimension variables in blood glucose meter design. The exterior shell is a series of quadratic Bézier curves with the control points along a rectangle around the design. The control points were each corner of the rectangle, indented down by the corner radius on each side, and the mid points along the top and bottom of the rectangle. ... 63
Figure 4.3. Screen capture of IGA interface with four blood glucose meters selected, denoted by a darker background. ... 65
Figure 4.4. Screen capture of IGA interface with four mobile phones selected, denoted by a darker background. ... 65
Figure 4.5. Select mobile phone designs, from left to right; the smallest final generation design selected, the mean design across all participants, and the largest design across all participants. ... 68
Figure 4.6. Select glucose meter designs, from left to right; the smallest final generation design selected, the mean design across all participants, and the largest design across all participants. ... 68
Figure 4.7. The four designs used in Experiment Two. Each uses the same independent variables. Blood glucose meter styles B and C use quadratic Bézier curves between the corners and points highlighted with stars to draw the shell around the screen and buttons. ... 71
Figure 4.8. Select designs, from left to right; the smallest selected final generation design, the mean design across all participants, and the largest selected design from all participants. ... 75
Figure 4.9. Comparison of glucose meter experience effect, which was not significant after correction. Mean final selected designs; left, designs from participants without blood glucose meter experience, right, designs from participants with blood glucose meter experience. ... 76
Figure 5.1. Independent dimension variables in mobile phone design ... 87
Figure 5.2. Screen capture of IGA interface with four mobile phones selected, denoted by a darker background. ... 90
Figure 5.3. The blood glucose meter independent variables. ... 95

LIST OF TABLES

Table 2.1. Intended audience of literature analyzed in the content analysis.................... 20
Table 2.2. Testable word groupings and frequencies within the literature, showing that the size of the device and components is the most frequently mentioned category of testable (non-abstract) terms in the literature... 20
Table 2.3. The mean relative size of mobile phones, screens, and buttons measured, showing how the various dimensions compare to each other in the existing phones found in the literature.. 21
Table 2.4. Paired T-Test *p*-values of the mean value of each variable for all phones and the selected phone(s) at the first and last iteration... 33
Table 2.5. Principal Component Analysis results; variance and the rotated component loadings of each variable. The values in the lower five rows show how much each variable contributes to the difference between participants. The higher the value, especially in the first component, the larger the effect. 34
Table 3.1. Prompts used in IGA software... 41
Table 3.2. Paired Wilcoxon Test comparing initial and final generation populations. 44
Table 3.3. *p*-values from pairwise Wilcoxon comparisons between prompt types shown in Figure 3.5 and Figure 3.6.. 45
Table 3.4. Man Whitney U Test comparing variance between local and remote groups. 47
Table 3.5. Prompts used for Experiment Two in IGA software...................................... 52
Table 3.6. *p*-values from pairwise Wilcoxon comparisons between touchscreen and non-touchscreen design trials shown in Figure 3.7... 54
Table 3.7. *p*-values from pairwise Wilcoxon comparisons between iPhone and non-iPhone owners, and touchscreen and non-touchscreen phone owners. 55
Table 4.1. Prompts used in IGA software... 64
Table 4.2. Mean (SD) of aesthetic values from Experiment One.................................... 68
Table 4.3. Prompts used for Experiment Two in IGA software...................................... 72
Table 4.4. Uncorrected paired t-test results comparing groups by glucose meter experience. All other were not significantly different (at an α equal to 0.05), and all variables here had *p*-values above 0.05 after being corrected for multiple comparisons. 73
Table 4.5. Dunnett's pairwise comparisons of the difference between the mobile phone prompt and each type of blood glucose meter design corrected for multiple comparisons. A higher *p*-value indicates a stronger correlation between mobile phone values and blood glucose meter values for that variable. (Only tests with a *p*-value α above 0.05 are included here.) 74
Table 4.6. Range of aesthetic values from Experiment Two, value (SD). 76
Table 5.1. Paired t-test *p*-values comparing the first and last generations, and the difference between IGAs alone and IGAs with GAs as well...................................... 91

Table 5.2. Mean aesthetic score (and SD) by trial type on a 0 to 100 scale, with 0 being extremely aesthetically unpleasing, 100 being extremely aesthetically pleasing. Mean (and SD) ergonomic score by trial type 92
Table 5.3. Mean (and SD) values of selected designs in the last generation by algorithm type. 92
Table 5.4. Paired t-test *p*-values comparing the first and last generations, and the difference between IGAs alone and IGAs with GAs as well. 96
Table 5.5. Mean aesthetic score (and SD) by trial type on a 0 to 100 scale, with 0 being extremely aesthetically unpleasing, 100 being extremely aesthetically pleasing. Mean (and SD) ergonomic score by trial type 97
Table 5.6. Mean (and SD) values of selected designs in the last generation by algorithm type. Pairs that were statistically significantly different ($\alpha = 0.0083$) are underlined. 97
Table 6.1. Chapter Four outcomes highlighting similarities and differences across domains 105
Table 6.2. Research outcomes relating to aesthetics. 106
Table 6.3. Research outcomes relating to physical ergonomics. 107
Table 6.4. Research outcomes relating to Aesthetic Ergonomic Research Methods 108

LIST OF EQUATIONS

Equation 2.1. Method of mating for each gene (independent variable) for each design in the next generation. 28
Equation 5.1. Method for calculating device length for device i, i=1,…,16.. The vertical button count was driven by device type (five vertical buttons on the mobile phone design). 88
Equation 5.2. Method for calculating keypad width for device i, i=1,…,16.. The horizontal button count was driven by device design (three horizontal buttons for the mobile phones designed). 88
Equation 5.3. Device Width calculation for device i, i=1,…,16.. Using the Screen Width variable and the calculated Keypad Width value for a device, the Device Width was equivalent to the larger of the two. 88
Equation 5.4. Length Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Device Length and Maximum Device Length are the smallest and largest device lengths for the population respectively. This scoring method gives the longest device a score of 10, and the shortest a score of 0. 88
Equation 5.5. Screen Area Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Screen Area and Maximum Screen area are the smallest and largest products of the screen width and height for the population respectively. This scoring method gives the largest screen a score of 10, and the smallest screen a score of 0. 88
Equation 5.6. Radius Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Device Radius and Maximum Device Radius are the smallest and largest device radius values for the population respectively. This scoring method gives the largest radius a score of 10, and the smallest radius a score of 0. 89
Equation 5.7. Device Width Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. The Minimum Width and Maximum Width are the smallest and largest device width values for the population respectively. This scoring method gives devices with a width at the mean of widths a score of 10, and devices at either the maximum or minimum width a score of 0. 89
Equation 5.8. Overall Genetic Algorithm fitness function combining the other equations above. 89

ABSTRACT

Aesthetic Ergonomics applies rigorous quantitative methodologies to the field of design to better understand aesthetic preference, potential tradeoffs between aesthetic and traditional ergonomic factors, and provide tools and methods that can be used in product design. This dissertation gives a better understanding of how users make aesthetic judgments within the domains of mobile phones and blood glucose meters, and demonstrates a tool and method for product designers.

Mobile phones and blood glucose meters are selected because they are common handheld digital input/output devices of functional and ergonomic consequence, whose users care immensely about their aesthetics. Interactive Genetic Algorithms (IGAs) and traditional Genetic Algorithms (GAs) are the primary research methodologies used in this dissertation. IGAs are a method of iteratively exploring a design space through a process of user-driven evolution that mimics natural selection. A GA, with a simplified physical ergonomics fitness function was used as a method of increasing the physical ergonomics of users' designs by combining it with an IGA.

Following the Dual Process Research Methodology, this dissertation uses top-down analysis to determine the aesthetic and ergonomic factors important in mobile device design, and then goes onto include bottom-up experiments to learn about aesthetic ergonomics, the use of genetic algorithms in aesthetic ergonomics, and mobile device design.

These studies found several ways that user aesthetic preference changes, specifically when: the user's goal(s) change, the way a device is used (touchscreen vs. non-touchscreen mobile phones) changes, and by experience. The research also found an aesthetic link between domains. IGAs combined with a physical ergonomic GA, to make mobile devices that were significantly more physically ergonomic, but not significantly less aesthetically pleasing to users, when compared to designs without a GA.

This dissertation has significant theoretical implications, including showing how users can hold multiple goals separate and combine them when designing devices. Practical implications include that the classic definitions of beauty, like the golden ratio, can serve as guides for designers, but do not apply perfectly to mobile devices.

Chapter 1

Introduction to Aesthetic and Ergonomic Design of Mobile Phones and Blood Glucose Meters

Customers often do not know or cannot express exactly what they want in a product (Liu, 2003; Noblet, 1993). This may be because they are trying to reconcile multiple goals, or it is hard to select from the infinite design options. As a result, product designers often rely on trial and error, or their own subjective judgments and heuristics instead of a data-driven approach to design (Otto & Wood, 2001). This use of designers' judgments and heuristics alone can lead to a time-consuming process and potentially non-optimal solutions. Reducing the required number of trials and prototypes to generate a better design will reduce the number of person-hours required in the design process, which will serve to lower costs, decrease lead-times, and make companies more competitive.

Aesthetic Ergonomics

The field of Aesthetic Ergonomics may, to the casual reader, lack an ostensible connection to the larger field of ergonomics referenced in its own title. The ergonomist or human factors professional will be aware, however, of the value of a pleasing experience to the quality of life of the user (Norman, 2002), as well as its benefit on their function or perceived function (Tractinsky, Katz, & Ikar, 2000). Aesthetics can play an important role for devices of such import to our productivity and safety. Two such devices are mobile phones and blood glucose meters. Both are devices where their functionality is important, but their usability is also heavily determined through aesthetic evaluation by their user.

Previously, Aesthetic Ergonomics has looked at variety of domains such as bottle design (Kelly, Maheut, Petiot, & Papalambros, 2011) and web page layouts (Bauerly & Liu, 2009). In the past, researchers have focused separately on aesthetics or physical ergonomics (usability). In doing so, designers and researchers removed the very systems nature from our field of systems engineering. Considering usability within aesthetic ergonomics will realize the goal of holistic design, creating designs that are more desirable and better for the user.

Dual Process Research Methodology

The dual process research methodology states that aesthetic design problems should be addressed from both a top-down process of discovering the major design factors by observation, and a bottom-up process of testing the specific effects of each factor on design, informing future top-down research (Liu, 2003). One of the primary goals of top-down research is to observe user choices to intuit the independent variables and understand the range of levels of those variables so that more detailed research can take place. Bottom-up research uses the findings from top-down research as a guide for variables and problems, but goes on to explicitly test theories and create experiments that have not naturally occurred. Using the dual process research methodology as a framework for this aesthetics research, this work is divided into two large parts: determining the key factors and levels to study, and then using the results to inform further investigations and future designs.

The initial focus in Chapter 2 is on completing the top-down portion. To make sure that the factors analyzed in the bottom-up research are the most relevant (in this case, affect aesthetic preference), they needed to be determined through a top-down processes. Once the top-down research is completed, it is validated using initial bottom-up experimentation (Chapter 2), and then more interesting aesthetic ergonomics concepts can be tested with a working bottom-up research tool (Chapters 3-5).

Modeling Aesthetics

Aesthetics are an important part of user satisfaction with individual models of mobile phones (Seva, Duh, & Helander, 2006). Aesthetics engender deep personal feelings that are contextually and temporally sensitive. User aesthetic preferences are intertwined with user experiences. The contemporary mobile phone aesthetic is vastly different from that of the first “brick” style or even “flip-style” mobile phones in the

1990’s. The differences reflect a change in user aesthetic preferences as well as changes in technological advancements that give designers new options.

The ways we use mobile phones have changed dramatically, affecting our design preferences as well. Mobile phones are now devices with a multitude of uses: voice communication, messaging, email, and web surfing; all of these uses play into our design preferences. With phones as a useful public tool, we have both a functional preference and a pure aesthetic preference (the difference being, our internal model of what we think represents a functional design, and what we think is beautiful). It is crucial for mobile phone designers to balance these individual goals and understand how they interact together.

Blood glucose meter aesthetics have also been an area of study (Licitra, 2011; Mastrangelo, 2011), and manufacturers have hired well known industrial design firms to design the meters to be more aesthetically pleasing as well. The field of aesthetic preference has been studied in detail, including aesthetic preference across cultures (Cons & Jenny, 1994), and guidelines for designing mobile devices and even mobile device interfaces (Gong & Tarasewich, 2004; Karlsson & Djabri, 2001). These guidelines however, are either lacking in their specificity or in their applicability to designing the blood glucose meters or mobile phones. Ergonomics and usability of mobile phones has also been tested (Plos & Buisine, 2006), as well as that of blood glucose meters (Story, Luce, & Rempel, 2009), but aesthetic ergonomics has not been tested in a way that can easily compared aesthetic preference across devices.

The design inspiration for new handheld blood glucose meters comes at least in part from mobile phones:

> *“the meter’s aesthetic design was inspired by consumer electronics—primarily because users’ expectations are set by all the other portable devices they carry with them, such as cell phones and MP3 players.” (Lloyd, 2010)*

The degree to which mobile phone preference shapes blood glucose meter preference is not clear, but a better understanding of this relationship can aide designers of mobile phones, blood glucose meters, and new products which don’t yet exist.

The ability to use computers to perform quantitative aesthetic analysis has allowed gestalt principles and the theories of artists like Paul Klee to be easily tested, giving rise to the field of quantitative aesthetics (Gero & Jun, 2006). Quantitative aesthetics applies mathematical analysis to a range of domains such as; image selection (Obrador, 2007), layout (Harrington, Naveda, Jones, Roetling, & Thakkar, 2004), and interface design (Ngo & Byrne, 2001) to test or use aesthetics theories, like those of Gestalt, and artist Paul Klee (Gero & Jun, 2006).

Several principles of aesthetics span across these fields, specifically the concepts of Balance, Unity, Rhythm, and Proportion. Balance refers to making the design components equally sized and spaced along any axis, or along all axis at once (Harrington, Naveda, Jones, Roetling, & Thakkar, 2004). Balance, sometimes referred to as visual weight, is quantified by calculating the distribution of design elements and their relative sizes. Balance applies to mobile device design in how the size of the screen can be offset by the keypad space (when the keypad is considered together as one visual element). The concept of Unity is closely related to Gestalt principles and seeks to calculate the degree to which design elements, such as the individual buttons, can be considered as one element. Unity measures inter-element spacing as well as element-border spacing to calculate the degree to which elements are closer to each other than to a border or other groups of elements. Rhythm, also referred to as movement, addresses apparent movement or changes in static designs. While Rhythm has a place in texture, assembly analysis, affordances, and interface design, it is not relevant here because of the static nature of the designs used and the two-dimensional testing interface (computer monitor). Proportion is one of the most well known aesthetics principles. Popular proportions include the square (1:1), the golden rectangle (1:1.618), and the square root of three (1:1.732) (Marcus, 1992). These models of aesthetics (Balance, Unity, and Proportion), are very useful models of aesthetics in this test of mobile phone preference and blood glucose meter preference.

Physical Ergonomics

Physical ergonomics in the handheld device arena is important. For example, mobile phones contribute to cubital tunnel syndrome (also known as "cell-phone elbow"), the second most pervasive peripheral nerve entrapment syndrome (Cutts, 2007). Cubital tunnel Syndrome is the result of compression of the ulnar nerve from prolonged elbow bending (often to put a mobile phone to your ear). Besides rest, cubital tunnel syndrome may require injections, physical therapy, or chiropractic care (Ashton, 2010). To reduce the likelihood of cubital tunnel syndrome, elbow flexion should be minimized to increase blood flow and reduce the length of the tendon, ideally below 80 degrees (Kahan, 2002). While a reduction below this threshold is likely not possible from changes to a mobile phone alone, it may be able to be mitigated by the use of hands-free devices or other interventions which reduce the amount of time spent holding a phone to the ear. Additionally, the risk of injury when holding a mobile phone to the ear for extended periods may be able to be decreased slightly by changes to the phone design investigated later.

Other areas of mobile device ergonomics have been investigated as well, grip span (Kong, Lee, Lowe, & Song, 2007), methods of determining thumb motion and finger force (Ong, 2009), finger abduction speed (Jonsson, Johnson, & Hagberg, 2007), and the effect of screen size on visibility (Hasegawa, Omori, Matsunuma, & Miyao, 2006). Even with this research, mobile devices have significant ergonomics problems beyond cubital tunnel syndrome. “Blackberry thumb,” the overuse of small handheld devices, can cause tendinitis in the thumb; pain and numbness in the thumb and joints of the hand (Gordon, 2008). Poor keypad layout and small keyboards found on mobile devices can slow down data entry (Balakrishnan, Yeow, & Ngo, 2005), cause pain for users with larger hands (Balakrishnan & Yeow, 2008), and their small size is often anthropometrically appropriate for children, not adults. This requires adults to adopt unnatural, uncomfortable postures to enter data (Croasmun, 2004). Interface legibility issues are also related to increased use, as the average users increasingly use their phone for messaging and email, their visual demands are increasing. Poor legibility can cause eyestrain and errors, and force users to adopt unnatural and uncomfortable positions. By modeling these two prevalent ergonomic issues, we can help designers better understand their tradeoff decisions in mobile device design.

Modeling Physical Ergonomics

Modeling physical ergonomics with the goal of design of handheld devices has been extremely limited for mobile phones and nonexistent for blood glucose meters. A kinematic model was created to recommend keypad dimensions for a new mobile phone design (Hirotaka, 2003).

A multi-objective physical ergonomics model for handheld mobile devices has not been proposed. Following the lead of another multi-objective physical ergonomics model (Brintrup, Ramsden, Takagi, & Tiwari, 2008), a generic model could provide guidance to designers across domains based on the common physical ergonomic problems of handheld devices. The perfect fidelity of such a model is not as important as having a model that would provide guidance towards an ergonomic solution that a designer could build upon.

Interactive Genetic Algorithms

The bottom-up component of the dual process research methodology consists of testing preference directly with users. Interactive Genetic Algorithms work in a similar manner to natural selection in the biological world. DNA from several sets of parents, who differ along a number of axes, are combined under imperfect conditions to create a generation of offspring that can be categorized as a combination of the parents' qualities along each axis, with some randomness. IGAs use a human subject as match-maker in selecting the most successful designs which will be used to create the next generation of designs. This next generation will have a mix of traits (levels of the independent variables) from the designs in the previous iteration. Interactive Genetic Algorithms were originally proposed in Dawkins (1996), and have been used more recently in aesthetic design such as (Bauerly & Liu, 2009; Kelly, 2008).

Using IGAs, the designs shown to the user constantly evolve towards something more preferred by the selector. To access the entire space as the designs evolve, every generation is exposed to some mutation, in which a small number of the variables' levels are changed randomly (similar to biological evolutionary mutation). To aid in exploring the design space, the designs that are not selected by the user are given a small chance to be one of the parents of a next generation of designs. Through this mating selection and mutation, it is possible to have several traits of a previously less appealing design or unseen design manifest themselves in future generations as a new, combined option that the selector may prefer.

IGA experiments parameterize the variables for a design into a "chromosome," with each "gene" representing an independent variable. For a particular design, each gene is set to a level of that gene's independent variable. For the entire population of designs each chromosome is a unique combination of levels of independent variables (genes). An example IGA evolution process is shown in Figure 1.1.

By controlling the number of designs at each iteration, the number of selections the human can make at each generation, the way in which mating combines the designs, the rate of random mutations and other parameters, an IGA can efficiently explore a multi-dimensional design space in a small number of iterations and find a preferred design. IGAs have previously been used for such purposes as tuning hearing aids (Durant, Wakefield, Van Tasell, & Rickert, 2004), and have recently been applied to testing aesthetic preference (Bauerly & Liu, 2009; Kelly, Papalambros, & Wakefield, 2006).

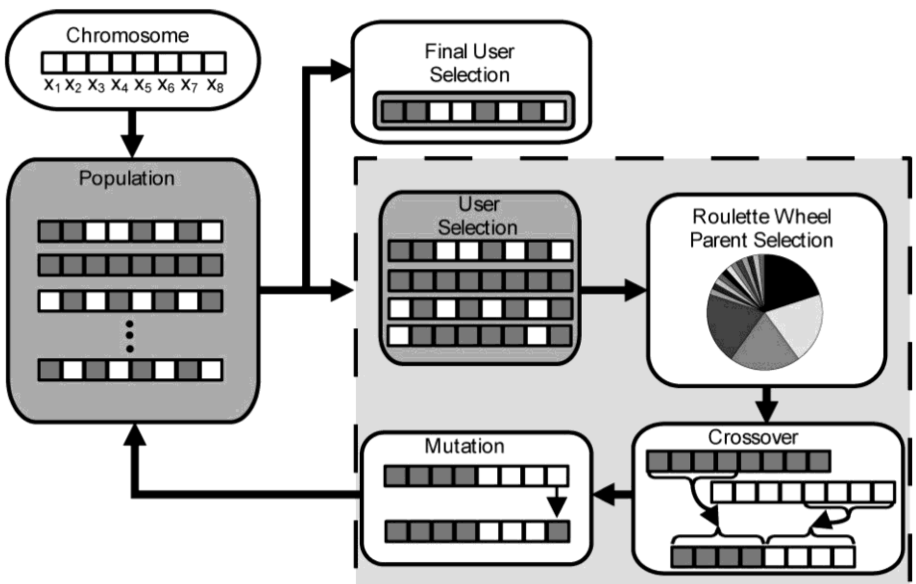


Figure 1.1. IGA Process diagram, showing how variables make a population of designs and user input as well as the genetic algorithm iteratively evolve that design (Kelly J. C., 2008).

By their very nature, human interactions with IGAs are mediated through use of a computer and are potentially affected by the computer environment. The user cannot physically touch the designs when using a computer-rendered IGA. The computer display may affect the size and proportions of designs, as well as the coloring, along with other aspects that may interact with the goals of the user. Keeping the display size constant is a method of controlling for these factors, which have been examined in previous studies.

With IGAs, the study participant performs the job of the fitness tester, using the criteria set by the experimenters. IGAs were selected as a method of testing because compared to other Design Of Experiments (DOE) techniques they offer several advantages: their ability to explore an entire design space quickly, and their inherent ability to change the resolution at which the IGA examines the problem.

Theoretically, a design space can be explored very quickly using Taguchi Robust Design, and other DOE methodologies that exploit orthogonal arrays. Utilizing these techniques, however, requires the use of very specific numbers of variables and levels of each variable to take advantage of the efficiency found in orthogonal arrays. This makes experiments using orthogonal arrays hard to set up, and usually has an impact on the design of your experiments by limiting the number of levels of each variable. Balanced Incomplete Block (BIB) ranking method studies (Liu, 2003), can also be used to explore a design space. Similar to the Taguchi Robust Design constraints, as the problem grows, and variables go from discrete to continuous, the number of blocks grows quickly. IGAs expose users to a wide variety of level combinations through the mutation and randomness of the IGA throughout the experiment. This allows participants to narrow their focus while still being exposed to new designs at the same time, leading to a faster overall solution (Kelly J. C., 2008).

As IGAs converge over the course of the experiment, the resolution, or magnitude of the differences between designs, becomes finer. At the start of the experiment, and through randomness and mutation, users are exposed to a wide variety of designs across the design space. As the experiment progresses however, the differences between the population of designs proposed to the user shrinks, and users are making minute choices between several barely distinguishable levels of continuous variables. This ability to test at both a high-level and very fine resolution, combined with IGAs speed make IGAs the ideal mathematical technique for this research.

Genetic Algorithms

A genetic Algorithm is a computer program that uses an equation to iteratively measure the success of variables in a design space as it explores that design space. Similar to Interactive Genetic Algorithms (IGAs), GAs set variables up as chromosomes and combine them to make designs. GAs and IGAs then iteratively evaluate designs against a fitness test, and select the best designs from a group to be used to make the next generation (subject to some mutation and combination of traits similar to evolution). GAs have been used to optimize vehicle suspension components (Alkhatib, Nakhaiejazar, & Golnaraghi, 2004), electromagnetics (Weile & Michielssen, 1997), and other fields. The inputs to a GA are the independent variables being changed in the design space. The dependent variable in a GA is the fitness test score, which determines whether or not a design is selected for use in the next round. The selections are based on the score of the combination of variables in the fitness test. The number of designs selected depends on the configuration of the GA. A Genetic Algorithm could be configured to use a certain number of designs for each iteration, all of the designs that exceed a certain threshold score, or some other selection criteria which selects from the final scores.

Genetic Algorithms have been combined with Interactive Genetic Algorithms (Brintrup, Ramsden, Takagi, & Tiwari, 2008), but have not been used to combine physical ergonomics and aesthetics. Using an IGA in combination with a GA would allow a balance between an aesthetic score and a quantitative genetic algorithm score, such as physical ergonomics. Combinations can be in parallel or asynchronous. In parallel combination, the GA and IGA run on the same set of designs, their selections are combined, and then iterated using the same method for determining the next generation. In an asynchronous combination, humans (via an IGA) and computers (via a GA) alternate selecting the designs that will generate the population for the other in the next iteration. While interesting, an asynchronous IGA GA combination may not converge as easily and may be difficult to analyze.

By combining an Interactive Genetic Algorithm with an ergonomic computer advisor (Genetic Algorithm), this work validates the ability of Interactive Genetic Algorithms to work with other agents as inputs simultaneously as with human operators. Part of the scientific significance of this work is that it will shed more light on the tradeoffs if any, that users make between their aesthetic preference and expert feedback, in this case in the form of an automated ergonomic consultant. Mobile phones and home-use medical devices were not what Plato and Kant were thinking off when they contemplated aesthetics. Using an IGA to determine aesthetic preference of mobile phones and home-use medical devices will allow us to test the classical theories of aesthetics in a new domain. This work extends the current research on aesthetic preference into a new domain.

From an applied standpoint, the developed tool will demonstrate the possibility of using Interactive Genetic Algorithms to reduce production time as a means of product design simulation, similar to other techniques. The research shows the ability of Interactive Genetic Algorithms to explore an entire design space from a high level and then with finer resolution in the areas of most interest. By developing this tool, designers will gain a proof of concept that will reduce the work required to develop a new design, and increase the designer's awareness of the tradeoffs that exist with their decisions. By reducing the required number of trials and prototypes, the tool will cut the design lead-time and reduce the amount of hours required to generate a new design. This will serve to lower costs and make companies more competitive. Through the education provided by the tool's automated feedback agent, in this case on ergonomics, designers will be better equipped to make tradeoff decisions. This tool will be a proof of concept that will demonstrate reduced development time with a more informed final design.

Research Domains

To investigate the aesthetics and physical ergonomics, the domains of mobile phones and Over-The-Counter (OTC) medical devices were chosen. These two fields were chosen because these consumer products are highly visible tools that people often invest in for both functional and aesthetic reasons and have significant ergonomic implications.

Consumers replace their mobile phones frequently, with many companies offering incentives to upgrade phones every year or two. Advances in mobile phone technology and changes in customer demands happen rapidly too, meaning that companies are continually trying to reduce mobile phone design lead-times to stay competitive. However, a poorly designed mobile phone may be difficult to use, and may cause accidents in time critical and multitask situations. This need to reduce lead-times, to consider aesthetic aspects, to maximize its use as a daily tool and as an aesthetic statement, and the importance of making the phones user-friendly and easy-to-use makes mobile phones an ideal choice for ergonomic aesthetic research.

Over-The-Counter medical devices, specifically blood glucose meters are a similar technology in that they are handheld digital input/output devices of immense functional import, with users who are incredibly sensitive to their aesthetics as well. Similar to mobile phones, the blood glucose monitoring market is continually working to avoid commoditization domestically as it grows internationally.

Mobile Phones

Mobile phones have become ubiquitous in modern society. With 4.6 billion mobile phone contracts worldwide, they are pervasive devices (ITU, 2010). Many people now spend the majority of their lives within just a few feet of their mobile phone(s). With new features and functions, mobile phones have become multi-use devices with aesthetic and functional components (McMullan & Richardson, 2006). Mobile phones today have changed the modality and frequency that people use to consume media, and communicate.

The prevalence and diversity of mobile phones have grown rapidly in the past decade. Today's mobile phone market can be broadly divided into two categories: touchscreen and non-touchscreen mobile phones. Within non-touchscreen mobile phones, there are "clamshell" (flip) types and "candy bar" (non-flip) mobile phones. Globally, the non-flip (non-touchscreen) type constitutes the majority, but touchscreen mobile phones represent the fastest growing segment.

Touchscreens are not new. They have been used in mobile phones since as early as 1992, when the IBM Simon Personal Communicator mobile phone was released. Although available now for almost two decades, touchscreen mobile phone usage has increased dramatically in the past few years. Arguably, this growth was led by the release of the Apple iPhone (Apple Inc.) in 2007. Since its introduction, the iPhone has been extremely popular, mobile phone, and is the most iconic touchscreen mobile phone to date. Touchscreen mobile phones are becoming more popular in "smartphones" and "non-smartphones" alike. Between August 2008 and August 2009 alone, touchscreen mobile phone usage grew 159% to 23.8 million users in the U.S. (comScore, Inc., 2009), or 10% of the total U.S. mobile phone market (comScore, Inc., 2010). This increase outpaced the 63% growth of smartphones of all types in the same year.

Mobile phones are selected by consumers for a myriad of reasons, and choosing a new mobile phone can be a daunting task. Phones compete across a number of factors, and are often chosen based on multiple goals, such as aesthetics, function, and battery life. Studies have shown that aesthetics has a major impact on mobile phone selection (Han, Kim, Yun, Hong, & Kim, 2004), but previous research does not test which components have the greatest impact on phone aesthetics. Additionally, previous research does not compare users' aesthetic preference with other user goals, or test preference across mobile phone types (i.e., touch screen vs. non-touch screen mobile phones). Both "candy bar" and touchscreen mobile phones are selected for testing in this dissertation because, respectively, they represent the most common and the fastest growing segments of the market.

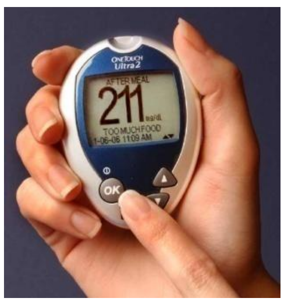
As mobile phones become more prevalent and replace more tools in our lives, our reliance on them grows. Consequently, the mobile phone market becomes more competitive. Increase in competition has forced the product lifecycle to become shorter, and the expected time to market of new designs must be more rapid. This shorter timeframe further increases the pressure on designers to understand quickly what users want.

Blood Glucose Meters

Home-use medical devices--medical devices that are used outside of a clinical setting--are often subsidized by health insurance providers, and users are allowed to pick whichever device they like the most. Blood glucose meters (also called glucometers) are the prototypical device in this rapidly growing market. Blood Glucose meters are provided to patients with diabetes to check their blood sugar levels. Patients of all ages and abilities often carry the device with them throughout their daily lives, using it two to ten times per day to calculate their blood sugar levels, often around meal times. Social stigma and self consciousness lead users to want a device that is visually innocuous, easy and comfortable to use. With consumers free to pick from a range of available interchangeable OTC devices (Figure 1.2), companies strive to differentiate their products with aesthetics, comfort, and ease of use. This drive for differentiation and rapid technology changes in the medical device industry lead manufacturers to have short lead times which depend on improving aesthetics and ergonomics, making glucose meters a perfect medical device which can be used to study aesthetic ergonomics research.

Accu-Chek Active Blood Glucose Meter

Prodigy Pocket Blood Glucose Meter



LifeScan OneTouch Ultra2 Meter

FreeStyle Flash

Figure 1.2. Sample of commercially available blood glucose meters.

Diabetes

Estimated to double globally from 171 million people in 2000 to 366 million people by 2030, diabetes (which is a leading cause of death in the U.S.), is adding to the amount of handheld personal technology many people use regularly (Wild, Roglic, Green, Sicree, & King, 2004). Indeed, the Blood Glucose Meter (the device used regularly by diabetics to measure the level of sugar in the blood) market more than doubled in size between 2000 and 2008 alone. With roughly 35 percent of the US Adult population classified as “prediabetic” (Centers for Disease Control and Prevention, 2011), the prevalence of blood glucose meters is likely to further grow.

Broadly defined, diabetes is a set of diseases where people cannot create enough of a chemical, insulin, which moves the glucose (sugar) we eat, from the blood stream in to fat, muscle, and liver cells where it can be used for energy, or these sites have a resistance to the insulin and don't use the insulin properly. Diabetics must be conscious of their blood glucose levels to avoid symptoms that include loss of vision, amputation, and even death. Starting around 1980, diabetics were able to test their blood glucose themselves using handheld blood glucose meters (Hughes, 2009). For many diabetics, blood glucose meters represent the best (or only) tool they have to regularly check their blood glucose level. This testing can be as infrequent as once a day at home, to as frequent as several times before and after meals in public. To test blood glucose levels, most blood glucose meters require patients to prick themselves with a lancing device, and put a small amount of blood on a single-use "test strip" which have chemicals that react to the blood allowing for measurement. The blood glucose meter uses these chemicals to calculate the blood sugar level and report it on a screen. Blood glucose meters come in a wide variety of shapes and sizes, from credit card or pen sized to devices larger than large conventional mobile phones. The patient or a caregiver must navigate a number of steps to prepare the digital device, take a reading, and in some cases store it for later retrieval or download. A detailed task analysis may be found in Rogers, et al. (2001). Additionally, many blood glucose meters have additional features, such as reading retrieval, backlights, or even connect to or contain digital data planners.

Medical device makers sell blood glucose meters and the disposable test strips. Most blood glucose makers subsidise the cost of the meters, making money instead on the recurring sale of single-use test strips. Even though diabetes rates are high and climbing domestically and increasing dramatically internationally, competition within the blood glucose meter space is intense. Commoditization, pressure from insurance agencies, competition from house labels, and discount mail-order type purchases have reduced profits slowing market growth even as the number of patients requiring monitoring globally soars. Technological advances have introduced bloodless blood glucose meters and subcutaneous continuous blood glucose monitoring, but they are still not commonly used. Regardless of the type of technology, designers are under significant pressure to design blood glucose meters that are safe, easy to use, attractive, and technologically advanced, while competing to differentiate themselves, keep costs low, and provide value to patients and healthcare professionals (Hughes, 2009).

References

Alkhatib, R., Nakhaiejazar, G., & Golnaraghi, M. (2004). Optimal design of passive linear suspension using genetic algorithm. Journal of Sound and Vibration , 275, 665--691.

Ashton, J. (2010, October 13). "Cell Phone Elbow": New Sign Of Times. (CBS, Interviewer) CBS.

Balakrishnan, V., & P. Yeow, P. H. (2008). Hand Anthropometry and SMS Satisfaction. Journal of Applied Sciences , 8, 816--822.

Balakrishnan, V., Yeow, P. H., & Ngo, D. C. (2005). An Investigation on the Ergonomic Problems of Using Mobile Phones to Send SMS. Proceedings of the Ergonomics Society's Annual Conference , 195-199.

Bauerly, M., & Liu, Y. (2009). Evaluation and Improvement of Interface Aesthetics with an Interactive Genetic Algorithm (IGA). International Journal of Human Computer Interaction , 25 (2), 155-166.

Brintrup, A. M., Ramsden, J., Takagi, H., & Tiwari, A. (2008). Ergonomic Chair Design by Fusing Qualitative and Quantitative Criteria Using Interactive Genetic Algorithms. Evolutionary Computation, IEEE Transactions on , 12, 343 -354.

Centers for Disease Control and Prevention. (2011). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

comScore, Inc. (2010, February 8). comScore Reports December 2009 U.S. Mobile Subscriber Market Share. Press Release . Reston, VA, USA.

comScore, Inc. (2009, November 3). Touchscreen Mobile Phone Adoption Grows at Blistering pace in U.S. During Past Year. Press Release . Reston, VA, USA.

Cons, N., & Jenny, J. (1994). Comparing perceptions of dental aesthetics in the USA with those in eleven ethnic groups. Int Dent J. , 44 (6), 489-94.

Croasmun, J. (2004, July 16). Are Ergonomists Really Consulted in Mobile Phone Design? Ergonomics Today . ErgoWeb.

Cutts, S. (2007). Cubital tunnel syndrome. Postgraduate Medical Journal , 83, 28--31.

Dawkins, R. (1996). *The blind watchmaker : why the evidence of evolution reveals a universe without design.* Norton.

Durant, E. A., Wakefield, G. H., Van Tasell, D. J., & Rickert, M. E. (2004). Efficient perceptual tuning of hearing aids with genetic algorithms. IEEE Transactions on Speech and Audio Processing , 12 (2), 144-155.

Gero, J. S., & Jun, H. J. (2006). Visual Semantics Emergence to Support Creative Designing: a Computational View. Australia: University of Sydney.

Gong, J., & Tarasewich, P. (2004). Guidelines for handheld mobile device interface design. 2004 DSI Annual Meeting (pp. 3751-3756). In Proceedings of the 2004 DSI Annual Meeting.

Gordon, S. (2008). Beware the 'Blackberry Thumb' New Repetitive Stress Injury Caused by Typing only with thumbs on hand-held devices. U.S. News & World Report.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S. W., & Kim, J. (2004). Identifying Mobile Phone Design Features Critical to User Satisfaction. Human Factors and Ergonomics in Manufacturing , 14, 15-29.

Harrington, S. J., Naveda, J. F., Jones, R. P., Roetling, P., & Thakkar, N. (2004). Aesthetic Measures for Automated Document Layout. ACM Symposium on Document Engineering (pp. 109-111). Milwaukee, Wisconsin, USA: ACM.

Hasegawa, S., Omori, M., Matsunuma, S., & Miyao, M. (2006). Aging effects on the visibility of graphic text on mobile phones. Gerontechnology , 4.

Hirotaka, N. (2003). Reassessing current cell phone designs: using thumb input effectively. (pp. 938--939). ACM.

Hughes, M. D. (2009). The Business of Self-Monitoring of Blood Glucose: A Market Profile. J Diabetes Sci Technol , 3 (5), 1219-1223.

ITU. (2010). Measuring the Information Society. International Telecommunication Union. Geneva, Switzerland: ITU.

Jonsson, P., Johnson, P. W., & Hagberg, M. (2007). Accuracy and feasibility of using an electrogoniometer for measuring simple thumb movements. Ergonomics , 50 (5), 647-659.

Kahan, N. J. (2002). Technique Addresses Computer-Related RSI. Occupational Medicine Clinical Care Update , 8 (17).

Karlsson, P., & Djabri, F. (2001). Analogue Styled User Interfaces: An Exemplified Set of Principles Intended to Improve Aesthetic Qualities in Use. Mobile HCI 2001: Third International Workshop on Human-Computer Interaction with Mobile Devices.

Kelly, J. C. (2008). Interactive Genetic Algorithms for Shape Preference Assessment in Engineering Design. Ann Arbor: University of Michigan Press.

Kelly, J. C., Maheut, P., Petiot, J.-F., & Papalambros, P. Y. (2011). Incorporating user shape preference in engineering design optimization. Journal of Engineering Design , 22, 627--650.

Kelly, J. C., Papalambros, P., & Wakefield, G. (2006). The Development of a Tool for the Preference Assessment of the Visual Aesthetics of an Object Using Interactive Genetic Algorithms. 9th Generative Art Conference. Milan, Italy.

Kong, Y.-K., Lee, S.-J., Lowe, B. D., & Song, S. (2007). Evaluation of various handle grip spans for optimizing finger specific force based on the users hand sizes. Human Factors and Ergonomics Society Annual Meeting Proceedings , 51, 884--888.

Licitra, E. (2011, April 29). Course ID 2012: Glucometer. Retrieved June 23, 2011, from Georgia Tech College of Architecture: http://www.id.gatech.edu/node/145686

Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. Ergonomics , 46, 1273-1292.

Lloyd, L. (2010, April). Survival of the Smartest. MD+DI; Medical Device and Diagnostic Industry , 32 (4).

Marcus, A. (1992). Graphic Design for Electronic Documents and User Interfaces. New York: ACM Press.

Mastrangelo, L. (2011, May 1). Course ID 2012: Sentinel Blood Glucose Monitor - Form and Function Project. Retrieved June 23, 2011, from Georgia Tech College of Architecture: http://www.id.gatech.edu/node/146492

McMullan, J., & Richardson, I. (2006). The Mobile Phone: a hybrid multi-platform medium. Proceedings of the 3rd Australasian conference on Interactive entertainment. Australia: Murdoch University.

Ngo, D. C., & Byrne, J. G. (2001). Another Look at a Model for Evaluating Interface Aesthetics. Int. J. Appl. Math. Comput. Sci. , 11 (2), 515-535.

Noblet, J. (1993). Industrial Design: Reflections of a Century. Paris: Abbeville Press, Inc.

Norman, D. (2002). Emotion & design: attractive things work better. interactions , 9, 36--42.

Obrador, P. (2007). Content Selection Based on Compositional Image Quality. IS&T/SPIE 19th Annual Symposium on Electronic Imaging. San Jose, California: SPIE.

Ong, F. R. (2009). Thumb Motion and Typing Forces during Text Messaging on a Mobile Phone. 23, pp. 2095-2098.

Otto, K. N., & Wood, K. L. (2001). Product Design: Techniques in Reverse Engineering and New Product Development. Prentice Hall.

Plos, O., & Buisine, S. (2006). Universal design for mobile phones: a case study. Proceedings of the ACM Conference on Human Factors in Computing Systems (pp. 1229-34). Montréal, Québec, Canada: ACM.

Rogers, W. A., Mykityshyn, A. L., Campbell, R. H., & Fisk, A. D. (2001, Winter). Analysis of a Simple Medical Device. Ergonomics in Design , 9 (1), pp. 6-14.

Seva, R. R., Duh, H., & Helander, M. G. (2006). Factors Affecting Emotional Responses in Mobile Phone Design Evaluation. Human Factors and Ergonomics Society 50th Annual Meeting, (pp. 2173-2177). San Francisco.

Story, M. F., Luce, A. C., & Rempel, D. M. (2009). Accessibility and Usability of Diabetes Management Devices for Users with Vision Disabilities. Human Factors and Ergonomics Society Annual Meeting Proceedings , 53 (11), 699-703(5).

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers , 13, 127--145.

Weile, D. S., & Michielssen, E. (1997). Genetic algorithm optimization applied to electromagnetics: a review. IEEE Transactions on Antennas and Propagation , 45, 343--353.

Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global Prevalence of Diabetes; Estimates for the year 2000 and projections for 2030. Diabetes Care , 27 (5), 1047-1053.

Chapter 2

Content Analysis, Anthropometric Modeling, and Exploratory IGA study of Aesthetic Ergonomics

Introduction

In order to make sure that the IGA varies the factors with the largest effect on users' aesthetic preference and to make sure that the "ergonomic teammate," the fitness function that later will be used to increase the physical ergonomics of the device design varies the factors that are the most important, two studies have been completed. First, a top-down study "to establish a 'global' and quantitative view of the critical dimensions involved in a specific aesthetic response" (Liu, 2003) was performed on a variety of media that were related to the aesthetics of mobile phones. Secondly, a top-down study of the leading ergonomic factors in mobile phone design was performed. Thirdly, a web-based Interactive Genetic Algorithm Software has been adjusted to test the factors found important to aesthetics on mobile phones.

Study One: Content Analysis of Literature on Mobile Phone Design Features

This study sought to identify the independent variables that relate to the aesthetics of mobile phones that should be varied by the IGA software.

Method

With the participation of several University of Michigan Undergraduate Research Opportunities Program (UROP) students, we conducted a year-long, comprehensive "content analysis" investigation of mobile phone aesthetic design factors. A variety of sources were identified to find articles and text materials that would relate to the aesthetics of mobile phones. A wide variety of different media were examined in this content analysis to determine which factors are important to mobile phone aesthetics. The types of media were: magazine articles, newspaper articles and advertisements, internet mobile phone articles, internet and TV mobile phone advertisements, and journal articles that touched on mobile phone aesthetics. These media were selected because they represent the various sources of information about mobile phones that are used by mobile phone designers and mobile phone users.

<u>*Procedure*</u>

Each of the articles and text materials was read by a trained content analysis research team member. Researchers were trained through readings on content analysis methods, such as the dual-process research methodology (Liu, 2003), and their work was checked by a Ph.D. student. Researchers recorded aesthetics-related words and recorded what other aesthetics-related words were mentioned in the same sentence or paragraph. They also coded the words by terminology class (size, color, feature, etc.); target audience (young readers, adult readers, males, females, academics); and by usage (e.g., the term ***small*** could refer to the screen size or overall phone size, and could be used as a positive or negative comment). Finally, the research team members measured the images of phones in the literature to catalogue the relative sizes of the screen, keys, and overall phone dimensions. The relative sizes were recorded to use later as reference for the size ranges of independent variables.

Results

The completed content analysis included an investigation of 118 articles or other pieces of media. Each of the 118 articles and other media were analyzed for words related to aesthetics and preference. Table 2.1 shows the articles and other media by their intended audience. Some of the articles were coded into multiple categories, because a magazine might be categorized as targeting both the youth and the male readers. For example, an article from the magazine Seventeen would be coded as both youth and female audiences, while the New York Times would be coded as an adult, but not gender specific audience, as it is not directly advertised towards a particular gender).

Table 2.1. Intended audience of literature analyzed in the content analysis.

	Articles	Words
Youth	43	284
Adult	39	17
Male	11	38
Female	11	18
Academic	33	934
All	118	1261

**Some articles fell into multiple categories*

Some of the words were more easily tested than others. After coding the words into testable and un-testable (abstract) categories, 213 testable words were categorized into seven categories (the majority of the words were un-testable). Words included in the testable categories include words like slim (thickness category), rounded (shape category), shiny (color category), or small / large (size category). The testable categories and the frequency of the words in them are in Table 2.2. Within the testable categories, size and color (color also referred to sheen as well as color) appeared significantly more often than others such as symmetry or features (such as slide-out keyboard).

As size emerged as the most relevant factor, when available, the relative sizes of mobile phones compared to their buttons and screen sizes were also recorded. The values of these measurements are presented in Table 2.3. The relative sizes recorded would result in a mobile phone similar to that shown in Figure 2.1.

Table 2.2. Testable word groupings and frequencies within the literature, showing that the size of the device and components is the most frequently mentioned category of testable (non-abstract) terms in the literature.

Term	**Appearances in the literature**
Sizes	86
Color	55
Textures	26
Shape	19
Thickness	15
Features	7
Symmetry	5
Total	213

** Abstract words, such as "trendy" or "sharp" were not included here.*

Table 2.3. The mean relative size of mobile phones, screens, and buttons measured, showing how the various dimensions compare to each other in the existing phones found in the literature.

	Width	Height	*N*
Phone	3.98	8.23	*52*
Screen	3.28	4.11	*49*
Button	0.65	0.50	*43*

Figure 2.1. Mobile phone design using mean measured values reported in Table 2.3. This phone is sized to have the mean proportions of all of the phones measured in the literature.

Discussion

Some of the most notable findings of the content analysis are the importance users placed on size, color, and texture, and the prevalence of un-testable, abstract words. The results of this study are in line with the factors identified in Seva et al., (2006), which found that size of certain attributes had an effect on emotion. Similarly, Chuang, Chang, & Hsu, (2001), and Han, Kim, Yun, Hong, & Kim, (2004) identified size of the phone or individual components as an important factor in preference or satisfaction. The sizes displayed in Table 2.3 and Figure 2.1 include touchscreen and non-touchscreen phones. Including measurements from touchscreen phones, which were a small portion of phones measured, may have slightly increased the relative screen size in relation to the overall phone dimensions. This slight effect does not affect the main finding, that the size of phones and their components are the elements that should be studied further.

The high frequency of color and texture categories is interesting to note. Although distinct categories in this analysis, color and texture are intertwined in design. A variable for texture may change the color or sheen of a device, and different colors can change the perceived texture. This is especially true when exclusively viewed as an image or video, as was done here. If, in future tests, researchers seek to test several colors and textures they should be different levels of the same independent variable to avoid confounded results.

The prevalence of abstract or hard to define words such as “sleek” or “stylistic” was notable. A majority of the words that were identified as related to aesthetics were these abstract, un-testable words. The difference can quickly be judged by looking at the difference in total words between Table 2.1 and Table 2.2. This limited the amount of information we were able to gain from each article.

Study Two: Physical Ergonomic Modeling of Two Problems with Mobile Phones

Despite widespread use, a review of ergonomic literature suggests that many ergonomic issues have not been resolved in mobile phone design. The literature points to two major ergonomic issues related to mobile phone design: Cubital Tunnel Syndrome and interface legibility issues. Cubital Tunnel Syndrome (CTS) is caused by having the elbow flexed for an extended period of time, such as when holding a phone to your ear. CTS is the second most common upper extremity compressive neuropath following carpal tunnel syndrome, and leads to pain similar to Tennis Elbow (Darowish, Lawton, & Evans, 2009). The results of poor legibility manifest themselves in a number of ways; through errors, dissatisfaction, and unnatural postures during use.

Method

After identifying Cubital Tunnel Syndrome and interface legibility issues as the primary physical ergonomic problems, two hypotheses were generated: 1) phone length positively correlates to interior arm angle, which has been shown to be related to Cubital Tunnel Syndrome (Darowish, 2009), and 2) screen size positively correlates to the maximum legible distance of the screen.

To test the first hypothesis, a very simplified anthropometric model was built that, for the sake of ease of calculation, assumed that the bottom long edge of the phone was held parallel to the ground at the user’s ear, as shown in Figure 2.2. The model used anthropometric data (Ergonomic design for people at work, 1983), and assumed that each component was a two dimensional straight line hinged at the connection to the other two dimensional lines.

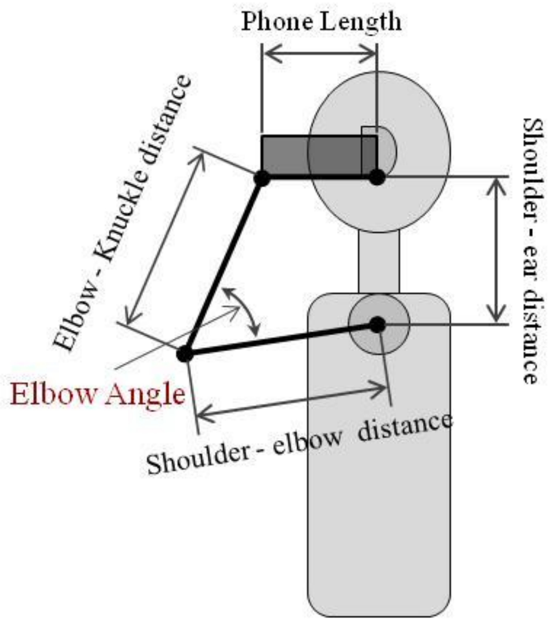


Figure 2.2. Simple anthropometric model relating phone length to elbow angle. This simplified model makes several assumptions about posture, and serves as an illustrative posture only.

The second hypothesis was tested by measuring the stroke width, character height, and spacing of two fonts (Times New Roman, and Arial), and calculating the maximum legible distance for three numbers of characters on the square screen (10, 75, and 180 characters on the screen) using the National Bureau of Standards (NBS) method. A Snellen ratio of 20/40 was used to compute the Effective Snellen Acuity. The character height and stroke width were calculated of each character density (number of characters on the phone screen), and the maximum distance was calculated. Characters had the same spacing and height:width ratio, but were sized so that the appropriate number of characters (10, 75, 180) would fit on the screen.

Results

The anthropometric model showed that phone length did affect interior elbow angle. A change in phone length of two inches could change interior arm angle up to ten degrees (Figure 2.3). A four-inch long mobile phone creates an arm angle of 39.4 degrees for the 50th percentile of the population. The 5th and 95th percentiles of the population have estimated average angles of 48.1 degrees and 37.9 degrees respectively. To reach an angle of 100 degrees, the phone would need to be approximately 18 inches long for the 50th percentile. It would need to be approximately 15 inches for the 5th percentile of the population and it would need to be approximately 20 inches long for the 95th percentile of the total population. This model neglects the wrist angle or the load on the shoulders because elbow angle is the primary concern.

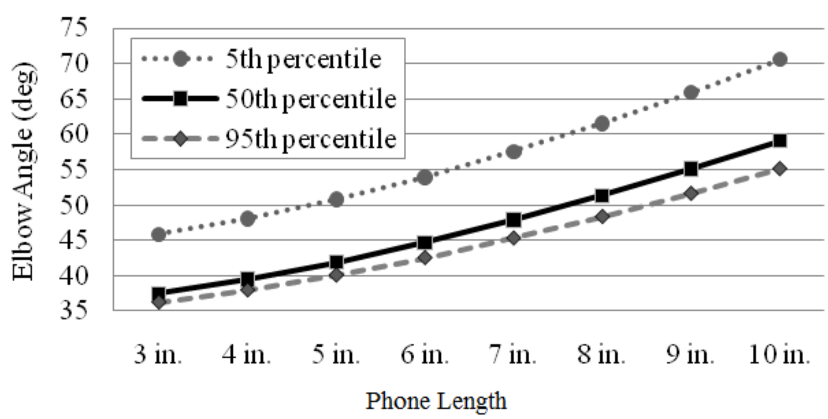


Figure 2.3. Elbow angle increase as a function of phone length.

Arial was found to be legible at a greater distance than Times New Roman using the NBS method. Ten characters using the default spacing of Arial font on a 4cm by 4cm screen would be legible from 153cm, while the same number of Times New Roman characters on the same screen would be legible at only 96cm. Figure 2.4 shows the mean of Arial and Times New Roman font maximum legibility distance based on screen size. The maximum legible distance was found to change as a function of screen size for all densities.

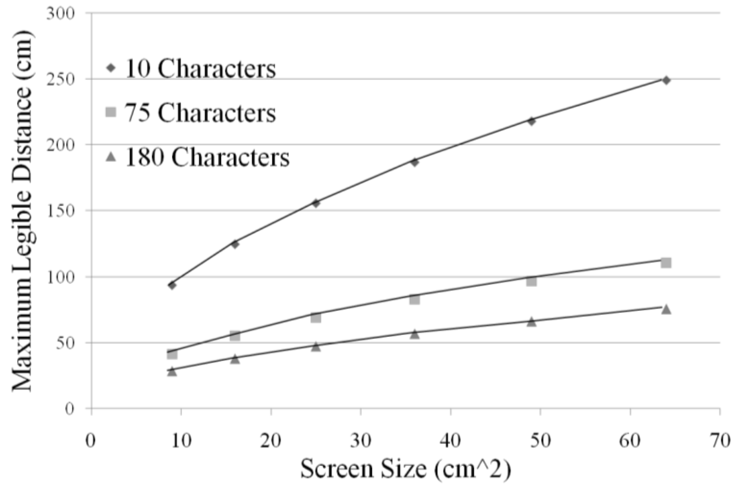


Figure 2.4. Mean of Arial and Times New Roman maximum legible distance for three characters densities on a square screen.

Discussion

The literature does not agree on the exact arm angle to eliminate the cause of cubital tunnel syndrome. The ulnar nerve stretches 5mm in length for every 45 degrees of flexion (Apfelberg & Larson, 1973). This implies that if the arm angle is 90 degrees, the ulnar nerve is stretched 10 mm from its normal position. Another article suggests that 45 degrees of flexion is considered the optimal position to decrease pressure on the ulnar nerve (Yamaguchi, Sweet, Bindra, & Gelberman, 1999). Another article suggested that the arm should be flexed no more than 30 degrees. The same article also stated that keeping the wrists straight and reducing pressure placed on a flexed elbow were two other good ways to help avoid cubital tunnel syndrome (Babski & Crumpton, 1997). A final article stated that elbow flexion should be 70-80 degrees from full extension (Kahan, 2002). Whichever you agree with, it is clear that mobile phone length can reduce the elbow flexion by several degrees while it is in this high-impact state of flexion. As this is a simple model focused on understanding CTS, this does not include study of the change in the moment arm based on the angle or distance away from the body the elbow is, or the lack of change on the muscle load. Both may affect comfort, and should be investigated elsewhere when developing a more robust physical ergonomics model.

The use of 10, 75, and 180 characters on the screen simulated a user reading a phone number, a menu screen, and a SMS (Short Message Service) message on their phone. From the legibility model, it is apparent that screen sizes are generally too small to be legible at great distances. It is possible that errors were made in assumptions for either model, but the trend, not the individual values are important to consider

These models linked specific phone attributes to two major ergonomic issues - one physical and one perceptual. These simple models will be used as inputs for the IGA tool in evaluating the joint effects of aesthetic and ergonomic factors, and can also serve as the basis of future work on the ergonomic inputs to mobile phones, or other mobile devices that have similar ergonomic issues. These models are not perfect or highly accurate, but are illustrative of the concepts and opportunities for physical ergonomics modeling of handheld devices. Future work can be done to explore other ergonomic relationships (e.g. button shape, thumb reach, button/body contrast), and can be done to more closely tie a reduction in risk or severity of Cubital Tunnel Syndrome with changes in arm angle on the scale shown here.

Study Three: Exploratory Study of Mobile Phone Aesthetics using an IGA

The goal of this study is to gather initial data that can be used to design further studies involving Interactive Genetic Algorithms in the handheld device domain. Additionally, this study seeks to understand users' preferences for the various combinations of design variables and levels, their discriminability between the different variables being changed.

By independently varying factors that have been identified as significantly affecting the overall aesthetic appeal of mobile phones, this study investigates users' preference for specific design variables as well as the relationships the variables have to each other. The results have implications for current designs by identifying factors which have the greatest impact on aesthetic appeal.

Method

Participants

Ten university engineering students, eight male, and two female participants were included in this study. The mean age was 24 years with a standard deviation of 2.7 years.

Variables

This study independently varies mobile phone horizontal and vertical button spacing and screen size as well as mobile phone corner radius for a simplified computer generated "candy bar" type phone. The button size on the phone was held constant at 2.1cm on the computer screen, with the mobile phone screen size, phone radius, screen radius, and spacing between the buttons changing between designs. The independent variables were horizontal and vertical spacing between buttons, screen width and height, and phone corner radius (the screen radius was half of the phone radius). The dimensions that varied are labeled in Figure 2.5. The spacing between the top row of buttons and the screen, and the minimum spacing between the phone edge and the screen and buttons were held constant. The dependent variable is participant preference, measured by selection at each iteration and collected via subjective questionnaire.

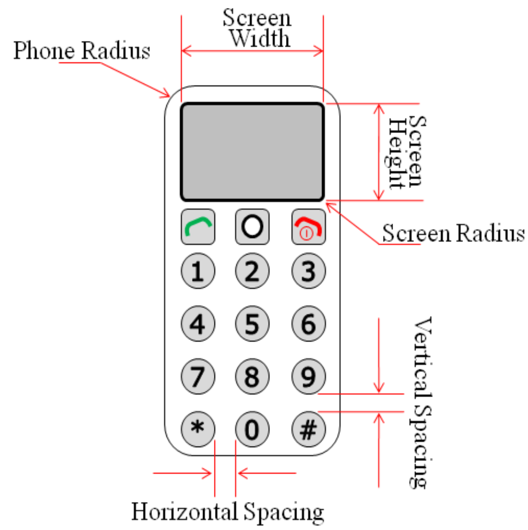


Figure 2.5. Independent variables being tested.

The values of the independent variables were set in PHP, a popular scripting language and read into Adobe Flash. After scaling onto the monitor, the actual size ranges of the independent variables were between 8.9 cm and 21.3 cm and between 6.4 cm and 15.9 cm for horizontal and vertical screen height respectively, between 0.4 cm and 3.2 cm for horizontal button spacing, between 0.4 cm and 2.1 cm for vertical button spacing, and between 0.4 cm and 4.3 cm for phone radius.

Interactive Genetic Algorithm Configuration

The Interactive Genetic Algorithm used in this experiment was based heavily upon the web-based IGA in Kelly (2008). The IGA started with 16 random designs in the first iteration. Parent selection happened through a roulette wheel function where every selected design got 20% of the roulette wheel (4 designs), and designs that were not selected split the remaining 20% of the roulette wheel (12 design, 1.66% each).

For each design in the next generation, the IGA selected parents from the current generation, mated the parents, and then subjected the entire population to random mutation (similar to Figure 1.1 in Chapter One). Parent selection was accomplished by using two random numbers that corresponded to locations on the roulette wheel to select parents for each design. Each design's chromosome consisted of the five independent variables shown in Figure 2.5. The order of the variables in the chromosome was horizontal button spacing, vertical button spacing, screen height, screen width, and radius. The IGA used this order in performing a mating of the two parents for each design. Mating was performed for each gene (independent variable) of each new design for the next generation using Equation 2.1.

$$OffspringGeneValue = Parent_1 GeneValue + RandomPercentage \times (Parent_2 GeneValue - Parent_1 GeneValue)$$

Equation 2.1. Method of mating for each gene (independent variable) for each design in the next generation.

After the entire next generation population was created using this method, the population was subjected to random mutation. Mutation was configured in the same manner as Kelly (2008), such that if mutation occurred for a new design, it would select one variable (gene) at random within that design, and change it to be a random value along the range for that variable. The rate of mutation was set so that there was a 12.5% chance that a design would have a mutation.

The IGA website ran locally (not over the internet) on a desktop PC, and a graphical display of candidate mobile phone designs were displayed in a 4x4 array

(Figure 2.6). Participants interacted with the PC through the use of a computer keyboard and mouse.

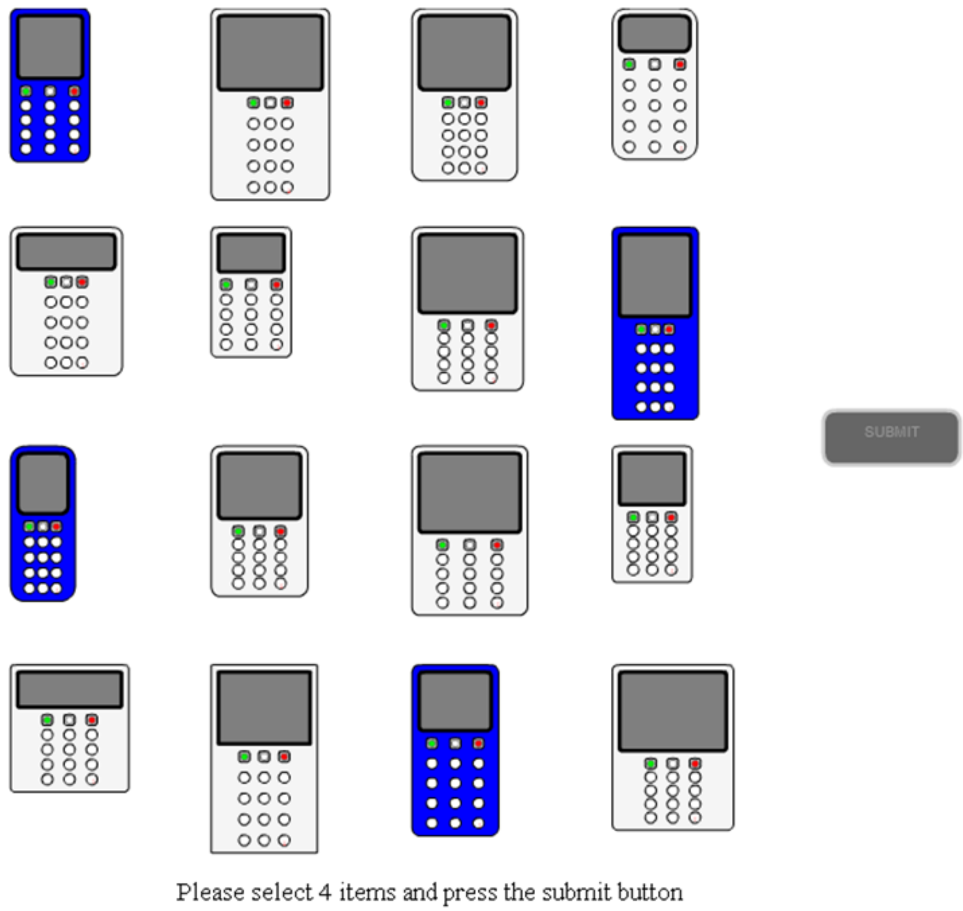


Figure 2.6. Screen capture of the software tool. The four phones with dark backgrounds around the screen and buttons indicate they are currently selected by the user.

<u>*Procedure*</u>

The experiment had four steps: qualification, introduction, testing, and feedback. Interested participants were qualified in person for the study. Anyone who expressed possibility of experiencing any discomfort using a computer for a period of an hour, was unfamiliar with computers, or did not own a mobile phone was excluded from participation in the study. After qualification, participants were given an overview of the study and the software as an introduction to the experiment. The overview showed participants how to use the software, explained the goals of the tool, and instructed participants to select the "most aesthetically pleasing" phones. After initial instruction, participants used the computer mouse to identify phone designs in the software tool. The software presented participants a quick overview of the buttons. After the overview, 16 phones were displayed on the screen; each generated using the variables outlined previously. Participants selected their top choice(s) for ten iterations. Participants selected four of the sixteen phones (Figure 2.6) for the first nine iterations, and their single favorite phone on the tenth iteration. After completing ten iterations with the software tool, participants completed a survey that included questions about their preference, the software's use, and their current mobile phone.

Data Analysis

The IGA data was saved to the local server using PHP to interpret the selections and access a MYSQL database. The level of each variable for all 16 phones was saved, as well as the designs selected at each iteration. SAS statistical software (SAS Institute Inc., Cary, NC) was used to perform paired T-Tests comparing the mean of the selected values on the first generation with the selected values on the last generation. Data was exported to SPSS (SPSS Inc., Chicago, IL) for Principal Component Analysis (PCA) to determine the relative importance of each of the variables.

Results

The results of the testing and feedback phases of the study showed that ten iterations of the IGA effectively identified design prototypes that appeared aesthetically pleasing to the participants. For every participant the IGA identified a local maximum at which the participant could no longer differentiate between designs. There was not a uniform preference for a level of each variable: designs varied greatly between participants (Figure 2.7), with some participants preferring smaller, skinnier phones, and others preferring larger phones with a larger button spacing as well.

A convergence towards a solution can be observed over successive iterations of the IGA, as the standard deviation (SD) decreased for each of the variables tested. The SD of the selected phones and the SD of all of the phones displayed decreased steadily across generations, leveling off at generation six. The population SD was consistently higher than the selected phones' SD. While not surprising, this shows that participants successfully "steered" the IGA towards designs they preferred. The changes in SD over the generations are shown in Figure 2.8 and Figure 2.9. When the SD of the population becomes approximately equal to the SD of the selected phones, the participants had converged on a solution for that variable. An example of this is in Figure 2.8, where the horizontal screen SD values for the population and selected values overlap at generation six.

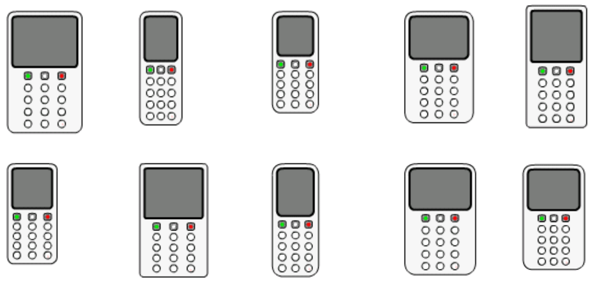


Figure 2.7. Final selected mobile phone designs, each design being the design selected on the tenth generation by a participation as the most aesthetically pleasing.

In the debriefing survey, all participants indicated a preference for the mobile phone they selected in the final iteration over previous phones presented to them throughout the experiment. Seven participants indicated interest in purchasing phones they had seen. Of the seven interested participants, five preferred their final selection for purchase, with the other two stating that any of the last few selections would have been fine. Although not part of the experimental protocol, many of the participants either commented in writing on the survey or verbally that by the final iterations the designs were hard to distinguish.

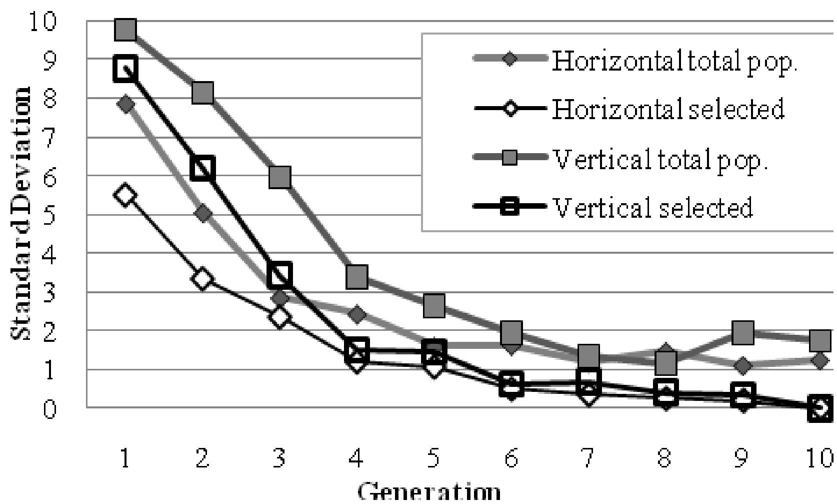


Figure 2.8. Standard Deviation (SD) of the horizontal and vertical screen dimensions by generation. The entire population's SD and the selected phone’s SD are shown.

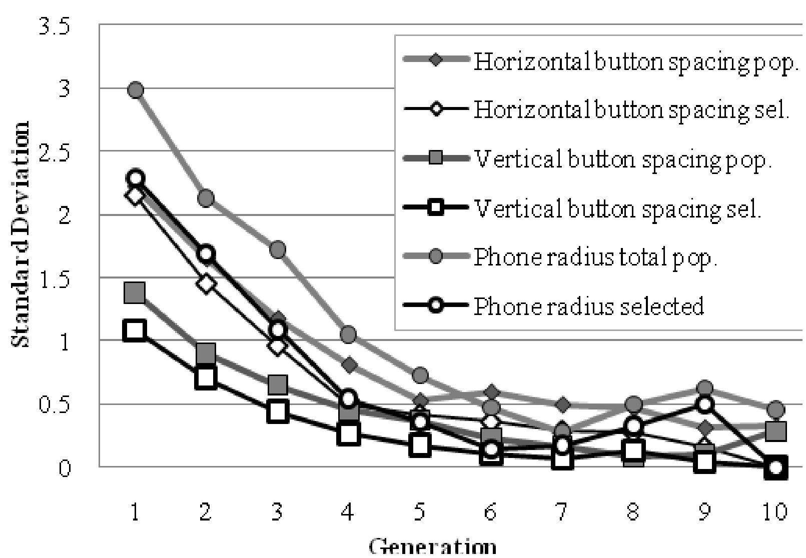


Figure 2.9. Entire and selected populations’ Standard Deviation (SD) of the horizontal and vertical button spacing as well as the phone corner radius by generation.

Participants felt most strongly about vertical screen size and button spacing differing from the mean, while caring the least about horizontal screen size and corner radius differing from the population mean. Comparing the phones shown to participants and the phones selected by participants in the first and last generations, there was a statistically significant difference between vertical screen size, vertical button spacing, and horizontal button spacing. This means that population shown to the participant changed throughout the test. Paired T-Test *p*-values comparing the starting and ending values are shown in Table 2.4.

Table 2.4. Paired T-Test *p*-values of the mean value of each variable for all phones and the selected phone(s) at the first and last iteration.

Variable	All phones	Selected phones
Horizontal Screen size	0.1561	0.1066
Vertical Screen size	**<0.0001***	**<0.0001***
Horizontal Button spacing	0.0517	**0.0327***
Vertical button spacing	**<0.0001***	**<0.0001***
Corner radius	0.1561	0.1066

* statistically significant at $\alpha = 0.05$

Participants' final phone selections were subjected to a Principal Component Analysis (PCA) using each of the five variables' levels for each participant. PCA is a regression methodology which identifies orthogonal (independent) regression models called components with each component attempting to explain as much of the variance as possible that was not already explained. In this data set, the first principal component explained 84.49% of the total variance in participant responses. The second and third principal components explained a further 10.97%, and 3.31% of the variance respectively. The remaining two dimensions explain only 1.22% of the total variance, and are not presented here. The percentage of the total variance and the individual component loadings are shown in Table 2.5. The table may be used to determine the relative weighting of the variables as they contribute to the variance amongst participants.

Table 2.5. Principal Component Analysis results; variance and the rotated component loadings of each variable. The values in the lower five rows show how much each variable contributes to the difference between participants. The higher the value, especially in the first component, the larger the effect.

	Component		
	1	2	3
Variance explained (%)	84.49%	10.97%	3.31%
Total variance (%)	84.49%	95.47%	98.78%
Horizontal screen size	0.985	0.13	0
Vertical screen size	0.134	-0.87	0.444
Horizontal button spacing	0.102	0	-0.397
Vertical button spacing	0	0	0.213
Corner radius	0	0.471	0.775

Figure 2.10 is the regression model, showing which variables have the largest effect on the first two principal components that are shown in Table 2.5. The first component, which explains most of the variance in the responses, is largely based on the horizontal size of the phone screen, and the second component ties the vertical screen size and radius together. The rotated component matrix graph shows the same variable relationships.

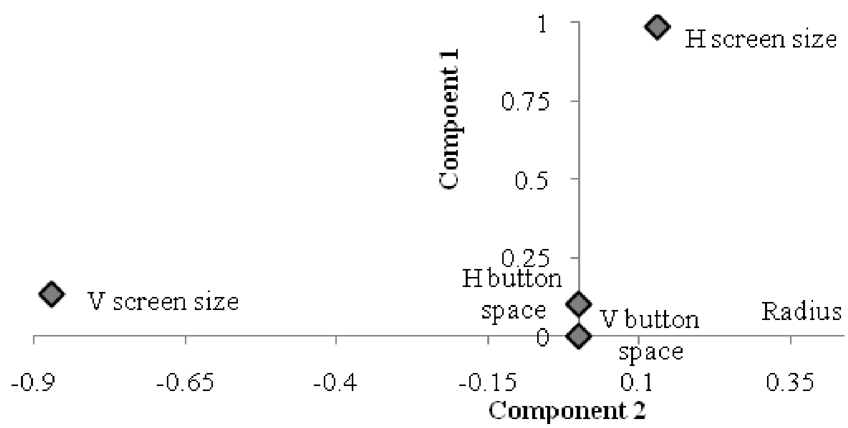


Figure 2.10. Variables plotted by loadings on components 1 and 2.

Discussion

This study sought to better understand user preference for the variables tested and their various combinations, gather initial data for further studies, and test the ability of the

Interactive Genetic Algorithm in determining user preference in this domain. Individual variable level preference was not the same for each participant, but anecdotally it appears there are trends which may lead to specific groupings of participants with a larger study sample size. The uneven number of males and females could have led to skewed results, something that could be fixed in a larger study.

When participants don't have a homogeneous preference, cluster analysis has been used by Kelly to segment users of a product similar to Kelly, Papalambros, & Wakefield (2006). Single or multiple products can then be brought to market and targeted specifically at these segments. Although individual variable preference was not determined, it is clear that horizontal screen size and horizontal button spacing were related, as were the vertical screen size and radius variables. The ability of participants to discriminate between IGA generated models leveled off around generation six, when the standard deviation of selected phones and the entire population of phones converged and stopped decreasing. This convergence indicates that the differences between designs became too subtle to distinguish, similar to Bauerly & Liu (2009), and Kelly, Papalambros, & Wakefield (2006).

Holding button diameter constant while varying screen size, the button spacing, and the phone radius worked well. It created a wide variety of plausible phone designs with independent variables. Within the ten iterations participants established a strong preference for their design, and could no longer further distinguish between the designs displayed. This demonstrates the ability of IGAs to effectively explore a large design space and quickly evolve a preferred design in the mobile phone domain. While other IGA configurations, such as different mating methods, or mutation algorithms may be more efficient or computationally compact, and should be tested in the future, the algorithm tested here was shown to work well enough, determining user preference within a small number of iterations, and is suitable for future use.

The finding that individual component size has an effect on preference of mobile phones strengthens previous findings about preference (previously in Chapter 2) in the mobile phone domain. The results also suggest implications for computer-generated interface designs and mobile phone designs, in that the relationships between specific size-related variables have an effect on overall aesthetic preference. This suggestion is similar to, but distinctly different than Seva, Duh, & Helander (2006), which found that size of certain attributes had an effect on emotion, and did not investigate preference. This work extends beyond the findings of Chuang, Chang, & Hsu (2001) and Han, Kim, Yun, Hong, & Kim (2004) who identified size of the phone or individual components as an important factor in preference or satisfaction but did not investigate the relationships between the variables. Combined, these previous studies and this work point to the importance of component size to user preference. The results do not investigate where our preference comes from, and if our internal representation of ourselves, or other internal models play into our sense of aesthetic preference.

The linking of phone radius and screen radius (screen radius was one-half of phone radius) likely led to the inverse relationship between the vertical screen size and radius in the PCA results. This study also shows that the use of Principal Component Analysis was very useful in understanding how variables interacted.

A potential limitation of this work is that a specific computer and monitor combination were used by all participants. This allowed the independent variables' values to be measured in terms of size on the screen, but also tied the results to a specific screen size. Future studies may consider using physical models or investigating the size of the screen (and therefore each phone) on preference, as larger or smaller representations may change the preferred relative sizes.

Further work is needed to understand the effect of prompt on preference and future studies may include other phone features as well. Modifying the prompt "select phones which you find most aesthetically appealing" versus "select phones which you find most functional" may affect users' internal fitness tests. Adding additional features, such as touch screen phones, "QWERTY" keyboards, or side buttons may give researchers and designers more insight into this important domain.

In conclusion, the results of this study show that specific relationships between the sizes of components that lead to mobile phone preference. This serves as further evidence for the usefulness of applying IGAs in the design process, and specifically identifies those factors which people deem most important to mobile device display. Future work will investigate additional factors to refine the IGA for mobile phone design as well as other applications, in order to ensure the robustness of this design tool.

References

Apfelberg, D. B., & Larson, S. J. (1973). Dynamic Anatomy of the Ulnar Nerve at the Elbow. Plastic & Reconstructive Surgery , 51 (1), 76-81.

Babski, K., & Crumpton, L. L. (1997). The use of ergonomic measures to prevent and control the occurrence of cumulative trauma disorders: carpal tunnel syndrome and cubital tunnel syndrome. Proceedings of the 1997 Sixteenth Southern Biomedical Engineering Conference, (pp. 407-410). Biloxi, MS, USA.

Bauerly, M., & Liu, Y. (2009). Evaluation and Improvement of Interface Aesthetics with an Interactive Genetic Algorithm (IGA). International Journal of Human Computer Interaction , 25 (2), 155-166.

Chuang, M. C., Chang, C. C., & Hsu, S. H. (2001). Perceptual Factors Underlying User Preferences Toward Product Form of Mobile Phones. International Journal of Industrial Ergonomics , 27, 247-258.

Darowish, M., Lawton, J. N., & Evans, P. J. (2009). What is cell phone elbow, and what should we tell our patients? Cleveland Clin. Jour. Med. , 76 (5), 306-308.

Ergonomic design for people at work (Vol. 1). (1983). Van Nostrand Reinhold.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S. W., & Kim, J. (2004). Identifying Mobile Phone Design Features Critical to User Satisfaction. Human Factors and Ergonomics in Manufacturing , 14, 15-29.

Kahan, N. J. (2002). Technique Addresses Computer-Related RSI. Occupational Medicine Clinical Care Update , 8 (17).

Kelly, J. C. (2008). Interactive Genetic Algorithms for Shape Preference Assessment in Engineering Design. Ann Arbor: University of Michigan Press.

Kelly, J. C., Papalambros, P., & Wakefield, G. (2006). The Development of a Tool for the Preference Assessment of the Visual Aesthetics of an Object Using Interactive Genetic Algorithms. 9th Generative Art Conference. Milan, Italy.

Liu, Y. (2003). The Aesthetic and Ethic Dimensions of Human Factors Design. Ergonomics , 1293-1305.

Seva, R. R., Duh, H., & Helander, M. G. (2006). Factors Affecting Emotional Responses in Mobile Phone Design Evaluation. Human Factors and Ergonomics Society 50th Annual Meeting, (pp. 2173-2177). San Francisco.

Yamaguchi, K., Sweet, F. A., Bindra, R., & Gelberman, R. H. (1999). The extraneural and intraneural arterial anatomy of the ulnar nerve at the elbow. Journal of Shoulder and Elbow Surgery / American Shoulder and Elbow Surgeons, 8, 17--21.

Chapter 3

Testing Aesthetic and Function Design Preference for Touchscreen and Non-Touchscreen Mobile Phones Using Interactive Genetic Algorithms

Introduction

Studies have shown that aesthetics has a large impact on mobile phone selection (Han, Kim, Yun, Hong, & Kim, 2004), specifically that the size and shape of the device and the individual components have the largest affect on mobile phone aesthetics (Seva, Duh, & Helander, 2006). Within size and shape, people vary in preference along several key dimensions. Prior work indicates that changes in horizontal and vertical screen dimensions account for a large portion of the variability between people's aesthetic preference (Chapter2).

The environment in which the IGA is used may also have effect on the outcomes. IGAs lend themselves to remote testing because of their automated nature, but this may alter the participant's results. Remote testing is very attractive in IGA studies because it significantly shortens the data collection time and decreases the workload of study participants and researchers, while potentially increasing the amount of data collected. This location effect is examined in great detail in Experiment One.

Goals

The goal of the two studies reported in this chapter is to build on previous IGA work on mobile phones that was purely aesthetics based. This chapter seeks to extend the literature by testing users' internal biases based on their specific goal or previous experiences. To do this, the studies test user ability to distinguish between multiple goals and also measure the effect of users' current mobile phone on their design preference. To make these comparisons, several models of aesthetic preference are tested against the data to see if and how they may be able to explain the different preferences between the designs.

Chapter Three Experiment One

This experiment builds on Chapter 2 in preference of handheld devices. The goals of Experiment One are to test user ability to discriminate between multiple design goals (function and aesthetics) and measure the effects of each goal on overall preference. In addition to these experimental goals, the effect on participant preference of a laboratory environment, compared to remote testing, when using an IGA, is tested.

Method

This experiment tests the effect of the goal that a user is given as a prompt on their preference. The measured user preferences are then tested against a number of the classical definitions of beauty to see if any of them accurately predicts preference for a particular goal. This study is performed locally and remotely, and the effect of the testing environment is investigated.

Participants

Twenty-one college students, 11 males and 10 females, participated in this study. The mean age was 24 years, with a standard deviation of 7.3 years. Participants were primarily engineering students (20 of 21). Participants used a wide variety of mobile phones; the most common was the LG EnV series (four participants, two models). The inclusion criteria were ownership or extensive use of a mobile phone, access to a computer with a high-speed internet connection and 19-inch to 21-inch monitor, and the ability to come to the research laboratory. Although not all participants actually performed the experiment in the laboratory, the ability to do so was an inclusion requirement for all participants. Ten participants performed the study in the lab, eleven performed in remote locations of the participants' choice.

Variables

There were seven independent variables: five within-subject independent variables built into the IGA software; as well as the within-subject variable of user goal (aesthetically pleasing, functional, or designs that are both aesthetically pleasing and functional); and a between-subject variable evaluating the effect of test setting (laboratory and non-laboratory settings). Between subjects, the study location was varied; subjects performed the study in the laboratory, or at a remote location of the participant's choosing which had a computer meeting the inclusion requirements.

The within-subject independent variables that were used to generate a computer-rendered mobile phone are similar to those used in the exploratory study explained in Chapter 2 and Nathan-Roberts & Liu (2010). They were: horizontal screen dimension, vertical screen dimension, the horizontal button spacing, vertical button spacing, and the "roundness" or radius (the radius of the outside of the mobile phone and the radius of the screen). The radius variable changed the exterior corner radius and the radius of the corners of the screen simultaneously. The screen corner radius was set to one-half of the exterior mobile phone corner radius. These variables can be seen in Figure 3.1.

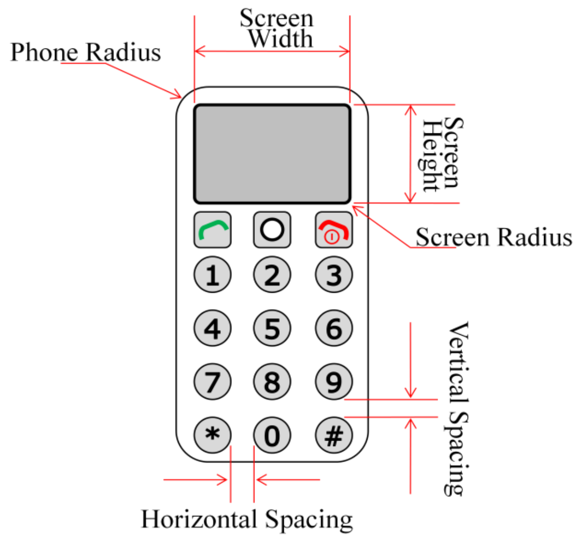


Figure 3.1. Independent dimension variables in mobile phone design

Within-subject, the prompt, or goal for participants in each of their seven trials, was changed. At each trial, participants were given one of four goals for their selection: practice (once), the most aesthetically pleasing design (twice), most functional design (twice), or both the most aesthetically pleasing and functional design (twice). The prompts are shown in Table 3.1.

Table 3.1. Prompts used in IGA software.

Goal	Prompt
Practice	"This is the first survey, and is a practice survey. Please experiment with the software by picking mobile phone designs that you like and see how the program responds by giving you new options in the next iteration."
Aesthetics	"This survey will be an aesthetics survey. Please select the phones that you find most aesthetically pleasing on every iteration."
Function	"This survey will be a function survey. Please select the phones that you find most functional on every iteration."
Aesthetics & function	"This survey will be a survey of both aesthetics and function. Please select the phones that you find most aesthetically pleasing and functional on every iteration."

The most relevant control variables within the study were: the button diameter, the number of iterations of each prompt in the study, the minimum spacing between the screen and outer edge, and the minimum spacing between the buttons and the outer edge.

The primary dependent variable of interest was the participant preference, measured by their selection at each iteration, and collected via subjective questionnaire.

Interactive Genetic Algorithm configuration

The Interactive Genetic Algorithm (IGA) in this study evolved designs over ten iterations. At each iteration sixteen mobile phone designs were shown. In the first nine iterations, participants selected their four favorite mobile phones; in the last iteration, they selected their single favorite. Figure 3.2 shows the user interface with four mobile phones selected.

The IGA had the same configuration as the IGA used in Study Three of Chapter Two, including the roulette wheel apportionment, parent mutation rate, and mutation method. It was coded in PHP, a popular scripting language, and read into Adobe Flash to render the mobile phones as a webpage. PHP recorded the displayed values and user feedback and saved it to MYSQL databases, an online database system. Variable ranges were set in PHP. The user interface webpage (Figure 3.2) displayed the participant's number, the current trial, and a link to proceed to the next trial or survey. A Flash webpage instructed users of their prompt (goal) before each trial.

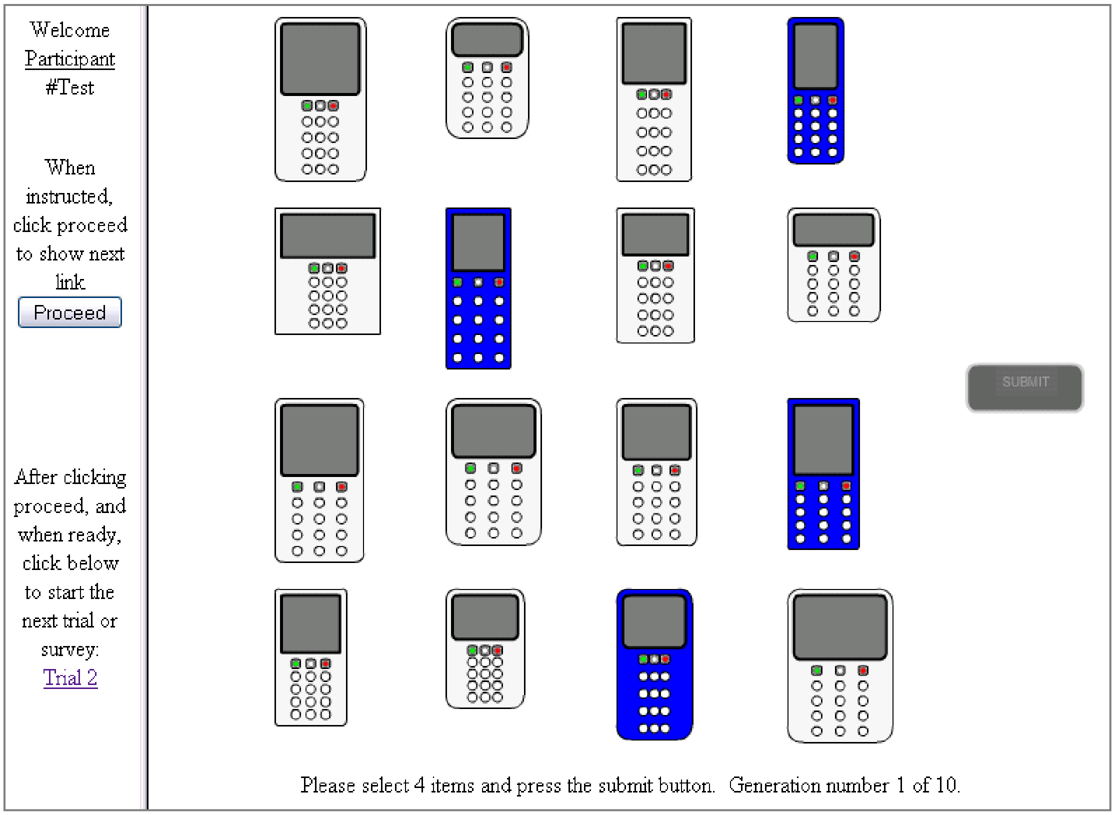


Figure 3.2. Screen capture of IGA interface with four mobile phones selected, shown with a darker phone background.

Procedure

Participants were directed to the experiment website via a recruitment email to university departments. Interested participants were evaluated via a screening form on the experimental website. Qualified participants visited the lab to fill out a consent form and receive instructions on using the experimental tool. Remote and local participants were given the same instruction sheet and overview of the software. Local participants completed the study using a laboratory computer with a high-speed internet connection and 19-inch LCD monitor. Remote participants completed the study outside of the lab on a computer meeting the inclusion requirements.

The experimental instructions include an overview of the study and the method to access the study online and use the software tool. Participants were told the most important thing for them to do was to “pick phones based ONLY on the prompt for that trial, and not based on previous prompts.”

Upon entering the study website, all participants were re-instructed about how to use the software, and the importance of following the prompt for their current trial only. Participants completed seven trials of the IGA software; one practice, and then six test trials. The six test trials were composed of two of each non-practice prompt organized in a random order. After each trial was a short questionnaire, and after all seven trials was a longer debriefing questionnaire.

Data analysis

After participants completed the study, results were pulled from the MYSQL database and saved locally to Excel (Microsoft Corporation). Statistical tests were performed in R (The R Project for Statistical Computing).

The normality of the distributions was tested with probability plots and Shapiro-Wilks W-Test. After determining that a portion of the results were non-parametric, Wilcoxon Signed Rank Test was used to test within-subject variables, and the Wilcoxon Rank Sum Test (equivalent to the Mann-Whitney U test) was used to test the location variable. Principal Component Analysis (PCA) was performed to determine which variables contributed most to the variance among participants.

Results

All twenty-one participants completed the study without reporting any major problems. All participants indicated that their final selections were “among the best presented” in that trial for the majority of trials (89% of trials, n=187 trials across all participants).

Several tests were performed to make sure that the IGAs’ algorithm was not deterministic, while still being able to converge and to exceed participant discriminability. Table 3.2 shows that for all variables, except radius, the first and last generations are significantly different, even when corrected for multiple tests. The standard deviation of the selected mobile phones and the population of mobile phones presented at each generation decreased steadily over each successive generation, until approximately generation six or seven. The population and selected mobile phone standard deviations are shown for each variable in Figure 3.3 and Figure 3.4. To test if the IGA exceeded participant discriminability, that is, if the differences were small enough between designs of a generation to be indistinguishable, participants were asked if the mobile phones ever started looking the same, and if so, at what generation that happened. Every participant stated that the mobile phones started looking the same during four or more of their trials. Overall, 89% of the post-trial questionnaires (n=187) said the designs started looking the same. The mean generation selected as the generation where mobile phones started looking the same was generation 6 (of the trials where participants said they started looking the same).

Table 3.2. Paired Wilcoxon Test comparing initial and final generation populations.

	Button W	Button V	Screen V	Screen W	Radius
p-value	<2.2e-16	0.001765	<2.2e-16	8.548e-07	0.3995

$\alpha = 0.01$ after Bonferroni correction.

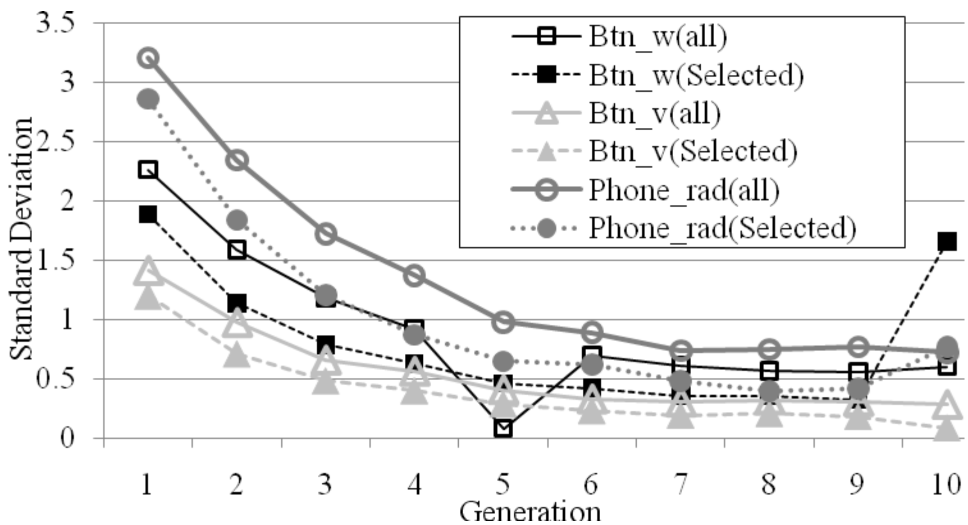


Figure 3.3. Standard deviation decrease by generation for the variables listed (selected designs and population as a whole).

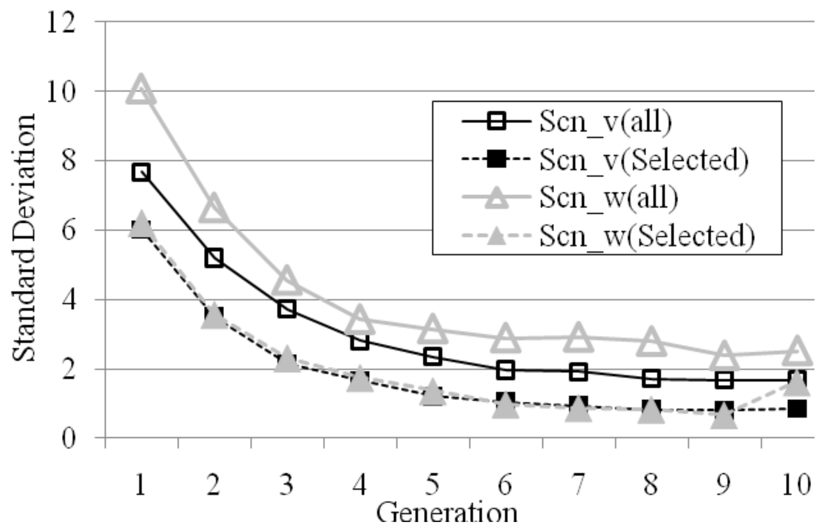


Figure 3.4. Standard deviation decrease by generation for the variables listed (selected designs and population as a whole).

Looking at the effect of prompt, the results show that each prompt led to a significantly different horizontal button spacing and screen width. Radius and screen height variables were not significantly different by prompt. The button spacing (width and vertical) and screen width were significantly different between the aesthetically pleasing and functional prompt types. Table 3.3 shows the *p*-values of each pairwise comparison. The mean for each variable in Adobe Flash pixel units, standard deviation, and lines showing statistically significant different pairwise comparisons are shown in Figure 3.5 and Figure 3.6 for the IGA.

Table 3.3. *p*-values from pairwise Wilcoxon comparisons between prompt types shown in Figure 3.5 and Figure 3.6.

	Button W	Button V	Screen V	Screen W	Radius
Aesthetics & Function	2.717 e-6	0.0005971	0.4028	2.078 e-11	0.004017
Aesthetics & Both	0.0006177	0.05743	0.6632	0.000872	0.4265
Function & Both	0.01256	0.1049	0.8924	4.599 e-5	0.08262

$\alpha = 0.0033$ (corrected with Bonferroni).

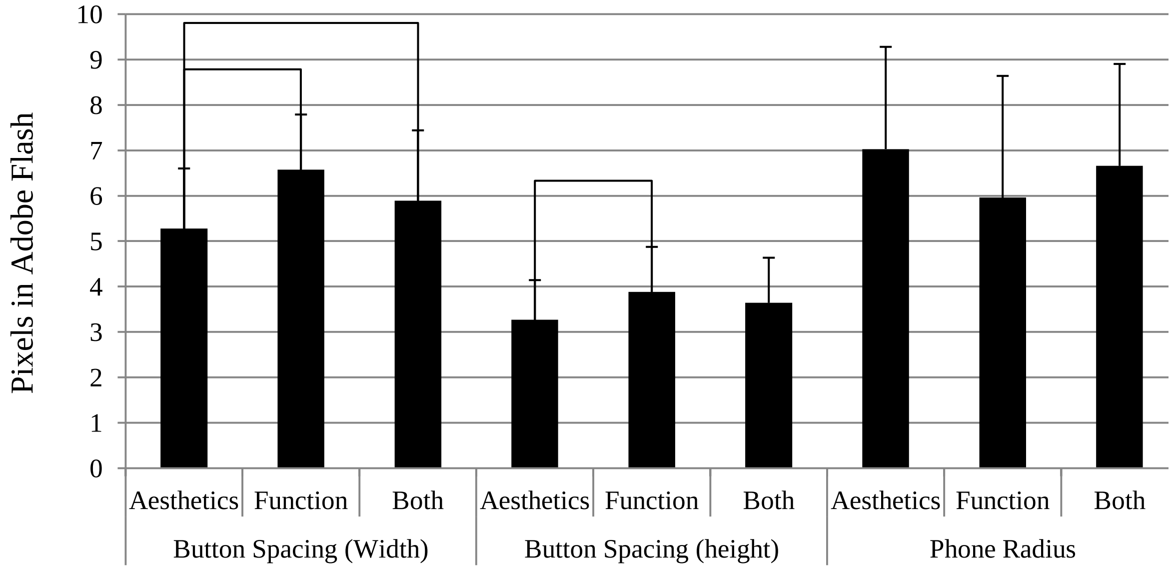


Figure 3.5. Graph showing mean values. Connecting lines indicate significantly different in pairwise comparisons at α = 0.0033 (corrected with Bonferroni). Error bars indicate standard deviation.

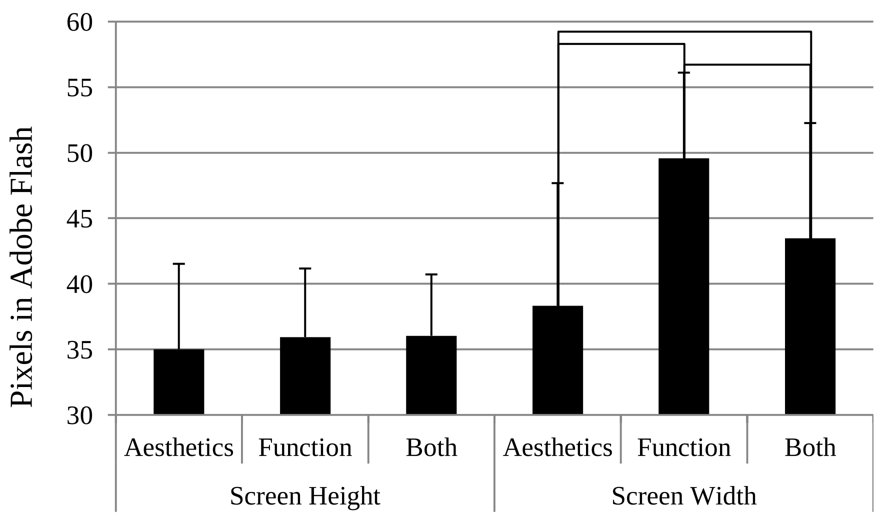


Figure 3.6. Graph showing mean values. Connecting lines indicate significantly different in pairwise comparisons at α = 0.0033 (corrected with Bonferroni). Error bars indicate standard deviation.

To test the effect of participant location, several statistical tests were performed. The most relevant one is the Mann Whitney U Test of variance, which did not reveal any difference in variables between groups. Table 3.4 compares the selected population variance (within participant) for the ninth generation (the final generation, with four designs selected) between local and remote participants. The lack of statistically significant *p*-values indicates that the populations did not have unequal variances, indicating that remote participants did not have more variance within their selections than local participants did.

Table 3.4. Man Whitney U Test comparing variance between local and remote groups.

	Button W	Button V	Screen V	Screen W	Radius
Aesthetics only	0.3494	0.3867	0.5573	0.0357	0.6047
Function only	0.4262	0.6539	0.08452	0.5573	0.4679
Both only	0.9725	0.1321	0.1517	0.061	0.3144

$\alpha = 0.0033$ after Bonferroni correction.

No discernible pattern emerged when testing the independent variables against three classic definitions of proportions: the golden ratio, square root of three, or 1:1 symmetry. There was no conclusive pairing relating to unity of either the keyboard, or the keyboard and screen relative to the overall frame. Additionally, the balance (visual weight of the keypad vs. screen size) was not significant by prompt. The results from Principal Component Analysis did not shed light on the relative importance of the independent variables because each component contributed between 12% and 40% of the total variance for each prompt type. This suggests the variance between participants was not a simple first order equation. The first component of each prompt contributed to 38-40% of the total variance, with the second component contributing to 21-23% of the variance.

To compare how exciting this experiment was to past IGA experiments, participants were asked to rank the experiment on a scale of one to seven, where one is “very boring,” and seven is “very stimulating”. Participants did not find this experiment very stimulating; the mean subject rating was 3.0 (SD 1.2).

Discussion

The study clearly demonstrates the ability of this IGA as a working tool for determining preference in the mobile phone domain. This is demonstrated by the results of testing whether or not preference is deterministic and the two tests of convergence, reduction in standard deviation of the population throughout the study, and participant feedback.

In testing whether or not the system is deterministic, a pairwise comparison of the first and last generation's means was used as a substitute of other tests here. While testing the difference between the first and last generation, as shown in Table 3.2, is not a true test of whether or not the system is deterministic, it tests whether or not the final results are significantly different than the mean of each variable's range. The foregoing is true because the initial conditions are random, so the mean of the first generation is the same as the mean of each independent variable's range. It would be a problem if the participants' desired state of each variable was also the mean of the range for that variable, but it is clear from our findings that this is not the case. As such, testing pairwise comparisons is a more efficient way to determine if an IGA's parameters are deterministic when the participant preference is not near the mean of the population than Monte Carlo simulation or time series analysis. These other methods would also be logical, and have been done elsewhere (Kelly, 2008).

Convergence is easily established by looking at the reduction in variance of both the population and selected group throughout the first half of the each trial. The standard deviation of the population is the mean standard deviation of the population that was shown to each participant at that generation, not of the entire potential population. The selected design standard deviation is similarly the average variance of all participants and trials for each generation. These serve as good indicators of convergence because they show that the population is becoming more homogeneous. This effect does have a limit, as the mutation inherent in the IGA will cause divergence in the population; increasing the variance, and potentially, to a lesser degree, to the selected populations' variance as well. Using variance as a measure of convergence is validated by using the subjective survey questionnaire data that are also presented.

Looking at the standard deviation plots and polling participants, we can see the point at which they exceed their ability to distinguish between most of the designs presented to them. Specifically, this is around 0.5 to 1 pixel in the Adobe Flash web page on the 19” to 21” monitors for the smaller variables, button spacing and radius; and around 2 pixels for the screen dimension variables. We can also see where participants are using the mutation to steer the designs by selecting mutations in later generations. This factor can be seen especially well in the selected variance of button width at generation ten.

The local/remote effect is an important consideration for future IGA work. The most important statistical test of the location variable is the Man Whitney U test of selected variance (between participants). This is because the largest difference between local and remote participants would be the focus with which they approach the experiment. By testing the selected population variance in the final generations of each trial, we see if the remote participants have a higher variance. A higher variance would suggest that they are not selecting their favorite designs as carefully, as they start to exceed their ability to discriminate between non-mutated designs.

The prompts chosen generated significantly different designs along several dimensions. The condition of both aesthetically pleasing and functional was different from the function and aesthetics prompts for the screen width, but not for the button vertical spacing (button V), and not significantly different from the function prompt for the button width. Looking at the *p*-values, it can be inferred that the “both” condition is closer to the functional design for horizontal and vertical button spacing, but not for the mobile phone radius.

The mobile phones designed for the aesthetically pleasing prompt were thinner and shorter because of the smaller vertical button spacing, and the mobile phones designed for the function prompt were wider (larger horizontal button spacing and screen width) and taller. It is important to remind the reader that these “functional” designs are being generated by the participants, and are not being tested for their functionality. That said, a taller mobile phone design, as was found with functional designs, has been shown to be related to anthropometric advantages that may reduce the likelihood of Cubital Tunnel Syndrome, also known as “Cellphone Elbow” (Chapter 2). In this study, untrained participants were able to create somewhat ergonomic or functional designs on their own.

A previous IGA study (Chapter 2) showed the value of Principal Component Analysis to determine how to stratify participants by taste. This study failed to find strong results in its PCA, but it should be noted that the aesthetically pleasing prompt found virtually the same order of variable loadings in the first component even though the amount of variance the first component accounted for was approximately half of previous work (Chapter 2). The increase in the effect of each variable on the total variance, and more nuanced interaction, may be caused by the participants becoming more sensitive to the independent variables and the IGA through their repeated exposure, something not previously tested. Sensitivity to the variables would also cause the variance reduction by generation shown in Figure 3.3 and Figure 3.4 to reduce more slowly, as participants can more carefully control the IGA. We see that the variance does level off later in this study when compared to the initial study (Chapter 2).

The use of the Bonferroni correction appreciably reduced the statistical power in this application as the familywise error rate may be much higher than the α prescribed by the Bonferroni correction. While other corrections, such as the Student Newman-Keuls post-hoc analysis may be slightly better, it would not noticeably change the results, and substantial literature points to the importance of using "good enough" methods (Kantowitz & Nathan-Roberts, 2009), which in this case Bonferroni corrections clearly are. Participants found this experiment slightly boring, rating it a 3.0 on a scale of 1 to 7, negligibly higher than a previous IGA study that used the same scale (Bauerly, 2007), which scored 2.8.

The experiment is meaningful because it tests participant location effect for IGA studies, and investigates our ability to test multiple design goals separately and together. Findings point to a lack of location effect in participants. Reduced experimenter workload, in-lab experiments being unnecessary, will allow for more IGA experiments and faster iterations. Additionally, this experiment demonstrates the ability of users to hold multiple design goals, such as aesthetics and function separately, and in combination, allowing investigators to study a variety of effects.

Leveraging the reduced workload from remote participants, it is logical to extend this work from testing user ability to hold different preferences based on usage goal of the same device, to test user ability to hold different preferences based on different usage mechanisms of the same device. To do this, Experiment One was followed by a second experiment testing the effect of touchscreen and non-touchscreen mobile phone technology on user design preference for the same task, dialing a phone number.

Chapter Three, Experiment Two

Experiment One tested the ability of users to use the goals of aesthetics and function separately or combined for the design of a mobile phone. Users can indeed hold these competing values in mind simultaneously and effectively use them separately or combined for design.

Experiment Two tests how a difference in function, if a mobile phone has a touchscreen or a physical keyboard, affects participants' aesthetic preference. In testing the effect of the function on aesthetics, we are testing an important interplay in device design; that is, how the function of a device drives our aesthetic for it.

Touchscreen and non-touchscreen mobile phones have different component sizes, averaging the two together can provide misleading results to designers for mean phone component sizes (Chapter 2). By keeping the device design variables the same, Experiment Two will be able to compare the effect of function on aesthetic preference.

Theoretically, if aesthetic preference is inherent in the user, preference would not change with the device technology (interaction method). Alternatively, if user aesthetic preference extends into the feel of the interaction, participant aesthetic preference would likely change with interface technology.

Method

Experiment Two further examines the effect of prompt by testing how user preference changes when the technology of a design is changed. The experiment also investigates the relationship between current mobile phone ownership and desired mobile phone designs.

<u>*Participants*</u>

Twenty college students, 10 males and 10 females participated in this study. The mean age was 22 years, with a standard deviation of 2.6 years. Participants owned one of three types of mobile phones: an iPhone (ten), another type of touchscreen mobile phone (three), or a non-touchscreen mobile phone (seven). The inclusion criteria were the same as Experiment One, without the requirement of participants being able to come to the laboratory. Replacing the criterion of coming to the laboratory, participants were required to reside at an U.S. address where payment could be mailed.

Variables

Within-subjects, Experiment Two used the independent variables previously shown in Figure 3.1 of Experiment One, although the user goal (prompts shown in Table 3.5) were changed. The between-subject independent variable was mobile phone type owned by the participant (an Apple iPhone, a non-iPhone mobile phone).

Table 3.5. Prompts used for Experiment Two in IGA software.

Goal	Prompt
Practice	"This is the first trial, and is a practice trial. The software tries to learn from your previous selections. Please experiment with the software by picking mobile phone designs that you like and see how the program responds using your selection to give you new options in the next step."
Touchscreen	"This trial will be a touchscreen trial. On every iteration please select the phone design(s) which you find most aesthetically pleasing as a touchscreen phone for dialing phone numbers."
Non-Touchscreen	"This trial will be a non-touchscreen trial. On every iteration please select the phone design(s) which you find most aesthetically pleasing as a non-touchscreen phone for dialing phone numbers."

Interactive Genetic Algorithm Configuration

In an effort to provide a more robust tool for experimentation, the mutation method was changed in Experiment Two. The previous mutation method (Chapter Two, Study Three and Chapter Three, Experiment One) selected designs using a mutation rate, and then randomly mutated one variable within selected design. The IGA was modified such that every variable may be mutated for every design. The mutation rate was now being applied to each variable of each design, instead of each design, so it was lowered from 12.5% to 5% to make sure that the designs would still converge. If a variable within a design was selected to mutate it was changed to a random value within the range for that variable.

Procedure

Participants were directed to the experiment website via a recruitment email to university department list-serves. Interested participants were qualified via a screening form on the experiment website. Qualified participants received information, filled out a consent form, received training, and completed the study remotely.

After completing a consent form, participants were given instructions about the tool and how to access it via email. Participants were instructed on the importance of making "their choices based ONLY on using the phone for number dialing, and to consider ONLY the current prompt." After accessing the study website, participants received further training on the software, and the importance of following the prompt for dialing a phone number only. The software used in Experiment Two evolved designs in the same manner as explained in Experiment One, and shown in Figure 3.1 and Figure 3.2. After training, participants completed a series of trials with a short questionnaire after each trial, and a longer debriefing survey after the last trial.

Participants completed nine trials; one practice, followed by eight alternating test trials (four non-touchscreen mobile phone selection, and four touchscreen mobile phone selection trials). Half of the participants started with a touchscreen selection trial after the practice, and half had a touchscreen trial as their second trial.

After all of the participants had completed the study, the data were analyzed in the same manner as in Experiment One.

Results

As a test for convergence, participants were again asked if the designs ever started to look the same, and if so, at what generation that happened. All participants indicated that the mobile phones started looking the same for at least seven of the nine different trials they completed. Overall, 97% of the post-trial questionnaires (n=187) reported the designs started looking the same. Of the trials where participants said the designs converged, the reported mean generation was 5.4 (SD 1.9). Although not shown here, similar to the first study, the standard deviation of the population and the standard deviation of the selections decreased steadily, leveling off between generations five to seven as well.

Designs varied by both design prompt and participant mobile phone ownership. Button horizontal spacing, screen height, and screen width were all statistically significantly different for each of the two feature sets (touchscreen and non-touchscreen designs). Table 3.6 shows the *p*-values for the pairwise comparisons between groups. Figure 3.7 shows the mean value for each independent variable, the standard deviations, and lines indicating significant pairwise comparisons.

Table 3.6. *p*-values from pairwise Wilcoxon comparisons between touchscreen and non-touchscreen design trials shown in Figure 3.7.

	Button W	Button V	Screen V	Screen W	Radius
p-value	0.003	0.039	0.005	<0.001	0.113

$\alpha = 0.01$ (corrected with Bonferroni).

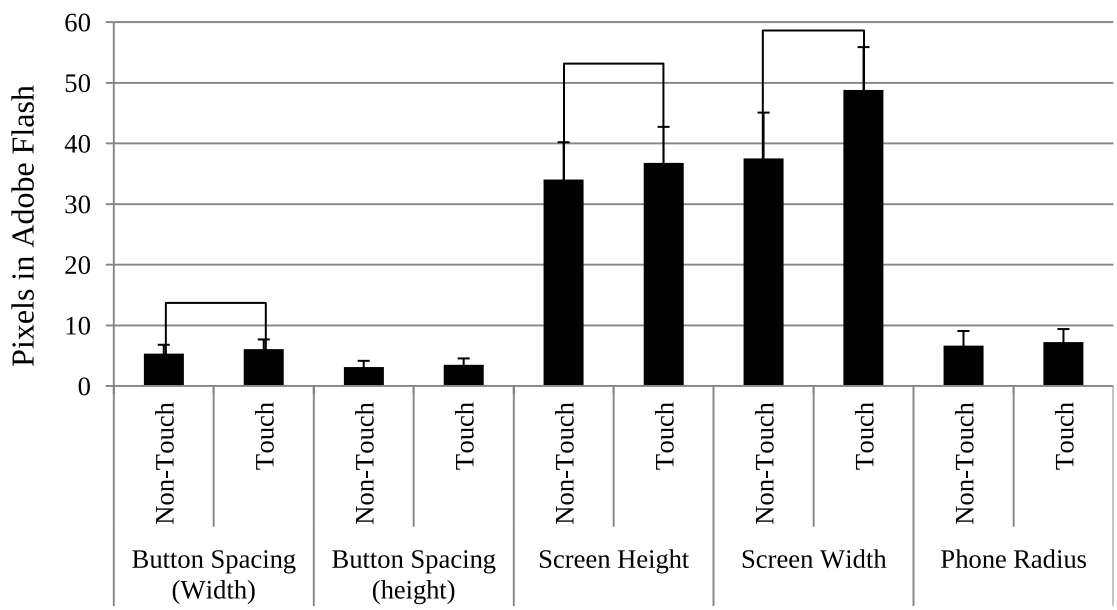


Figure 3.7. Graph showing mean values. Connecting lines indicate significantly different mean values in pairwise comparisons at $\alpha = 0.01$ (corrected with Bonferroni). Error bars indicate standard deviation.

Principal Component Analysis again failed to yield useful insight into the variables that cause the most variance amongst participants. The variance explained by each of the 5 components was 8.6% to 29.8% for touchscreen design loadings, and 11.4% to 36.6% for non-touchscreen design loadings.

Participant mobile phone ownership did have a significant effect on design preference, specifically the screen height variable. iPhone owner preference was compared to non-iPhone owner preference. Touchscreen owner, and non-touchscreen owner preferences were compared as well. Pairwise comparison results between iPhone, and non iPhone owners as well as touchscreen (iPhone, and non-iPhone touchscreen phone) owners, and non-touchscreen owners are shown in Table 3.7. Touchscreen owners who did not own iPhones had design preferences that more closely aligned with iPhone owners than non-touchscreen owners. Overall, touchscreen phone owners preferred a significantly taller phone screens. The mean design preference for each combination of prompt and ownership (touchscreen owner, and non-touchscreen owner) varied as well (Figure 3.8).

Table 3.7. *p*-values from pairwise Wilcoxon comparisons between iPhone and non-iPhone owners, and touchscreen and non-touchscreen phone owners.

	Button W	Button V	Screen V	Screen W	Radius
iPhone vs. non-iPhone users	0.357	0.055	0.006	0.070	0.819
Touchscreen vs. non-touchscreen users	0.097	0.026	0.002	0.086	0.198

$\alpha = 0.005$ (corrected with Bonferroni).

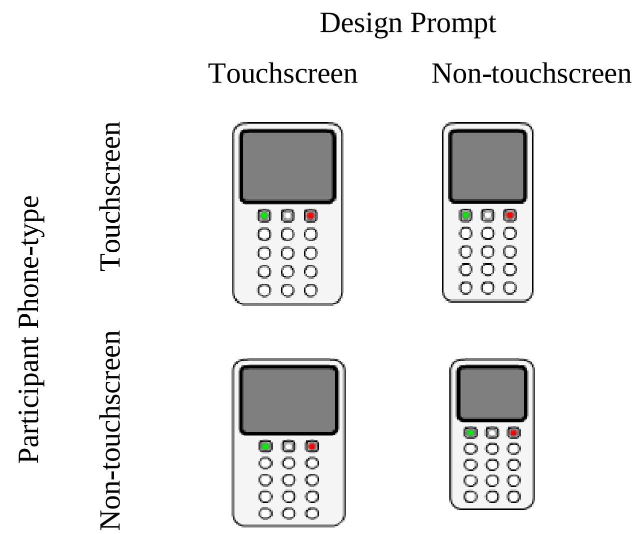


Figure 3.8. Mean design for each prompt separated by the type of phone owned by participant showing touchscreen phone owners preferred taller keypads and statistically significantly taller screens.

Discussion

Subjective and quantitative convergence data are presented to show that the change in mutation did not negatively affect the function of the IGA. The mean generation where participants rated designs as looking the same was actually lower than in Experiment One. This would suggest an increase in future mutation rates when applied to every variable to match the previous IGAs rate of convergence. Determining how participants differed through the use of Principal Component Analysis was again not possible.

While button width and screen width variables remained significantly different between prompts in Experiment Two, button vertical spacing was no longer significant. Yet screen height became significant. Overall, touchscreen mobile phone designs were wider and taller than non-touchscreen designs in Experiment Two. The mobile phone screen height variable, linked to the mobile phone screen size, was likely especially affected by the touchscreen prompt because touchscreen mobile phones are known for having larger screen areas.

Although other papers refer to the relative sizes of components (Dong & Ding, 2009), no studies to date have looked at the effect of touchscreen / non-touchscreen mobile phone design on aesthetic preference.

General Discussion

Overall these studies point to the ability of users to hold separate design preferences depending on their goal or the way a device is used. These studies also further strengthen the argument for using IGAs to determine aesthetic preference. The horizontal independent variables (horizontal button spacing & screen width) were affected in both studies, indicating that they may be more sensitive measures. Screen width, and horizontal button spacing to a lesser degree, have been shown to be a primary contributor to aesthetic preference between participants in mobile phone design (Chapter 2). The major difference between the studies was which vertical independent variable was significant.

In Experiment One, a significant pairwise comparison demonstrated that participants had different preferences depending on their goals. This is helpful because it can easily inform design. Designers can look at their use goal or target market (focused on aesthetics, or on highly functional, or a combination of both), and design accordingly.

In Experiment Two, however, significant pairwise comparisons reveal that participants had significantly different aesthetic preference between two devices that would be used for the same task. This outcome can be interpreted in two ways, first that participants could not focus only on mobile phone usage for dialing numbers as they were prompted, and instead imagined completing other functions with the mobile phones, leading to such dramatically different results. The is that the change in the physical interaction mechanism (touchscreen vs. physical buttons) contained inherent aesthetic considerations that changed the user aesthetic preference. The distinction between these interpretations is how well participants were able to design mobile phones following the prompt exactly, that is, for dialing only, instead of integrating their intrinsic knowledge of the other phone uses into the design. It is most likely that the significantly different values of the independent variables are a combination of both effects on the aesthetic evaluation: participants ability to strictly follow prompts, and also the appearance of touchscreen technology.

Here then surely, as discussed previously, the use of the strong Bonferroni correction of multiple tests may have inflated the rate of false negatives, or decreased the number of significantly different groups. The only variable even close to being significant would be vertical spacing of the buttons. It can be seen from the *p*-value that almost no relevant "weaker" correction would have recognized this, or the radius *p*-value as statistically significant.

In evaluating the IGA used in these studies, we find that this IGA worked well as a tool to determine design preference in novice users. This IGA has some advantages over others in the literature; for example, the web-based nature of the IGA allowed for easier online studies. The option of subjects engaging in remote testing greatly reduced the workload associated with in-lab participant studies during Experiment Two. Another advantage is the selection procedure of picking the "top four" is cognitively simpler than ranking procedures used by others (Cho, 2002).

The IGA used here can potentially be improved. Using experienced participants, or performing more practice trials prior to the test trials, may improve the ability of the IGA to record and understand the relative weightings of independent variables as they relate to preference. Increasing the mutation so the IGA exceeds user discriminability in a later generation than six may cause the studies to become more intrinsically interesting to the participants. More attention from participants, therefore, may provide more nuanced data about participant preference. The mutation parameter could also be improved using methods such as simulation annealing to inject mutations that would better explore the design space.

The current study builds on prior work establishing the importance of mobile phone size and shape (Chapter 2; Chuang, Chang, & Hsu, 2001; Han, Kim, Yun, Hong, & Kim, 2004; Seva, Duh, & Helander, 2006). The relationship between the size of attributes and emotion (Seva, Duh, & Helander, 2006) is tested in reverse, with the participants modifying the size of an attribute based on their emotional goal (preference for the design). This study serves as a dynamic test (and confirmation) of the findings of others (Chuang, Chang, & Hsu, 2001; Han, Kim, Yun, Hong, & Kim, 2004), who found the size of the mobile phone, or of an individual component of the mobile phone, affected satisfaction. The results here go further to relate the various components and identify their relative importance in various goals. While others have looked at mobile phone component size, they have not allowed participants to see components together (Dong & Ding, 2009), or they have taken an endogenous approach (Plos & Buisine, 2006), as opposed to the exogenous or "bottom up" approach outlined in (Liu, 2003) and used here.

With the rapid changes in mobile phone technology, the individual values and designs in this dissertation should be seen more as a snapshot of current technology, preferences, and design norms, and not as a recipe for future designers. While Hedonomics, Emotion, and Affective Design were not explicitly tested here, aesthetic preference, functional preference, and the technology types that were tested, are all inherently intertwined with the formerly mentioned areas of study.

A key limitation of this study is that the designs were not tangible; participants could not physically hold or interact with them. Had participants been able to physically interact with the devices, especially for a significant period of time, their ratings of function may have been different. As such, the function ratings serve as estimates by an untrained non-ergonomist population, and should not be used by designers as actual ergonomic or functional recommendations. For an introduction to handheld device ergonomics, the reader may be directed to (Balakrishnan & Yeow, 2008; Jonsson, Johnson, & Hagberg, 2007; Nathan-Roberts, Beeker, & Liu, 2009). Future work should take place in four areas: further testing of user ability to hold multiple separate goals, expansion to and comparison with other domains, further refining of the IGA, and additional research into handheld device physical ergonomics. This chapter demonstrates an important methodology for end users or designers with specific goals in mind, to contribute the design of devices. At the same time, user ability to hold and compare multiple goals using IGAs, or other similar methodologies, warrants further study.

While important, mobile phone research is not new. Testing IGAs in new domains, such as handheld medical devices, would be an instructive use of IGAs. Further research and improvements in the developing field of IGAs may allow for faster, easier, and more robust testing.

In conclusion, this chapter establishes IGAs as a valid design testing algorithm for personal devices, and as a tool to disambiguate and compare user goals. Further research in goal testing, other domains, and IGA improvements is needed.

References

Balakrishnan, V., & Yeow, P. H. (2008). Hand Anthropometry and SMS Satisfaction. Applied Sciences , 8 (5), 816-822.

Bauerly, M. (2007). Mathematical modeling and experimental study of interface aesthetics. Ph.D. dissertation, University of Michigan, United States -- Michigan: Dissertations & Theses @ CIC Institutions.(Publication No. AAT 3276095).

Cho, S. B. (2002). Towards creative evolutionary systems with interactive genetic algorithm. Applied Intelligence: the International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies , 16 (2), 129-38.

Chuang, M. C., Chang, C. C., & Hsu, S. H. (2001). Perceptual Factors Underlying User Preferences Toward Product Form of Mobile Phones. International Journal of Industrial Ergonomics , 27, 247-258.

Dong, W., & Ding, M. (2009). A Kansei-Based Method of Product Form Design and Its Application in Mobile Phone. Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on, 11-13, pp. 1-4. Wuhan, China.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S. W., & Kim, J. (2004). Identifying Mobile Phone Design Features Critical to User Satisfaction. Human Factors and Ergonomics in Manufacturing , 14, 15-29.

Jonsson, P., Johnson, P. W., & Hagberg, M. (2007). Accuracy and feasibility of using an electrogoniometer for measuring simple thumb movements. Ergonomics , 50 (5), 647-659.

Kantowitz, B. H., & Nathan-Roberts, D. (2009). Sources of Stimulus-Response Compatibility: Frames, Rules, and Response Tendencies. The Ergonomics Open Journal , 2, 163-169.

Kelly, J. C. (2008). Interactive Genetic Algorithms for Shape Preference Assessment in Engineering Design. Ann Arbor: University of Michigan Press.

Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. Ergonomics , 46, 1273-1292.

Nathan-Roberts, D., & Liu, Y. (2010). Investigation of Relative Mobile Phone Size Preference Using Interactive Genetic Algorithms. 54th Annual Meeting of the Human Factors and Ergonomics Society. San Francisco, California: HFES.

Nathan-Roberts, D., Beeker, A., & Liu, Y. (2009). Modeling Two Key Physical Ergonomic Problems with Mobile Phones. University of Michigan Engineering Graduate Symposium. Ann Arbor.

Plos, O., & Buisine, S. (2006). Universal design for mobile phones: a case study. Proceedings of the ACM Conference on Human Factors in Computing Systems (pp. 1229-34). Montréal, Québec, Canada: ACM.

Seva, R. R., Duh, H., & Helander, M. G. (2006). Factors Affecting Emotional Responses in Mobile Phone Design Evaluation. Human Factors and Ergonomics Society 50th Annual Meeting, (pp. 2173-2177). San Francisco.

Chapter 4

Comparison of Design Preferences for Mobile Phones and Blood Glucose Meters for Experienced and Non-User Groups

Introduction

Goals

The goals of these studies are to test the aesthetic preference relationship between two personal, public, digital input/output devices; quantify some basic aesthetic preferences for blood glucose meters, and test the IGA in a new domain.

Chapter Four, Experiment One

This experiment builds on prior work by the authors (Nathan-Roberts & Liu, 2010; Nathan-Roberts, Kelly, & Liu, 2011) in preference of handheld devices. The goal of the first experiment is to test how closely related user aesthetic preference is for two similar products; mobile phones, and handheld blood glucose meters. In addition to this goal, the question of user experience with the devices is raised, and preliminary analysis is performed.

Method

<u>*Participants*</u>

Twenty-two students, 12 males and 10 females participated in this study. The mean age was 22 years, with a standard deviation of 2.6 years. Participants were largely engineering students (20 of 22). Participants had a wide variety of modern mobile phones. The inclusion criteria were: ownership, or extensive use of a mobile phone; access to a computer with a high-speed internet connection, and 19 inch to 21 inch monitor; being free from pain related to several hours of continuous computer usage; normal or corrected to normal vision; and an U.S. address where payment could be mailed. Participants were also queried about their experience using blood glucose meters, eighteen participants (ten males, and eight females) had experience using a blood glucose meter, and four did not.

<u>*Variables*</u>

The independent variables are similar to those used in the exploratory study explained in Nathan-Roberts & Liu (2010), with the addition of the within-subject variable of testing a blood glucose meter, or a mobile phone, and the between-subject variables of blood glucose meter usage experience, and phone type used.

The within-subject independent variables that were used to generate a computer rendered mobile phone previously were used here again, and were also used to generate the blood glucose meter designs. The variables were the horizontal and vertical screen dimensions, the horizontal and vertical button spacing, the roundness (the radius of the outside of the mobile phone or blood glucose meter and radius of the screen). The radius variable varied the exterior radius and the screen radius simultaneously. For mobile phones, the screen radius was one half of the exterior radius value (Figure 4.1). The exterior shape used for the blood glucose meter, Figure 4.2, consists of a series of quadratic Bézier curves which were a function of the corner radius variable. The screen radius was one fourth the corner radius variable.

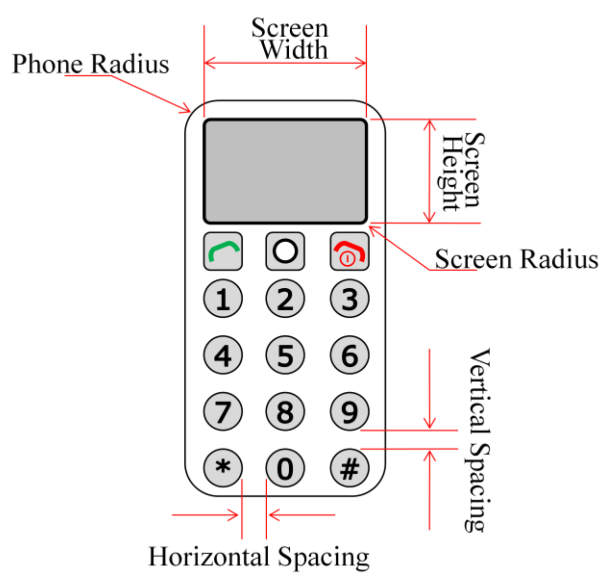


Figure 4.1. Independent dimension variables in mobile phone design.

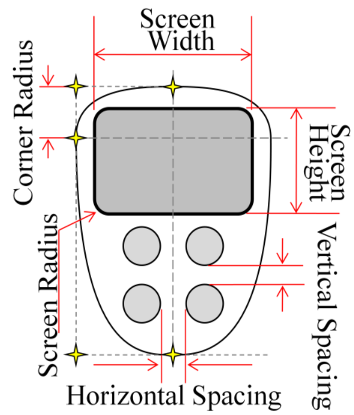


Figure 4.2. Independent dimension variables in blood glucose meter design. The exterior shell is a series of quadratic Bézier curves with the control points along a rectangle around the design. The control points were each corner of the rectangle, indented down by the corner radius on each side, and the mid points along the top and bottom of the rectangle.

Within-subject, the design alternated between designing a mobile phone or a blood glucose meter for each for each of their eight trials. At each trial participants were given one of three goals for their selections: practice (twice), the most aesthetically pleasing mobile phone design (three times), or most aesthetically pleasing blood glucose meter (three times). The exact prompts are shown in Table 4.1.

The most relevant control variables within the study were held constant across mobile phones and blood glucose meters; the button diameter, the minimum spacing between the screen and outer edge, the minimum spacing between the buttons and the outer edge, and the number of times participants experienced each prompt in the study.

The primary dependent variable of interest is the participant preference, measured by their selection at each iteration, and collected via subjective questionnaire. Additionally, the subjective questionnaire recorded measures of interest in this study and other subjective measures.

Table 4.1. Prompts used in IGA software.

Goal	Prompt
Practice	"This is a practice trial for glucose meters. The software tries to learn from your previous selections. Please experiment with the software by picking designs that you like and see how the program responds using your selection to give you new options in the next step."
Mobile phone	"This trial will be a phone trial. On every iteration please select the design(s) based ONLY on which you find MOST aesthetically pleasing."
Blood glucose meter	"This trial will be a glucose meter trial. On every iteration please select the design(s) based ONLY on which you find MOST aesthetically pleasing."

Interactive Genetic Algorithm configuration

The Interactive Genetic Algorithm (IGA) in this study had the same configuration as the IGA outlined in Study Two of Chapter Three on page 52, including a mutation function which had a lower mutation rate, but was applied to every variable on every design. The only difference was that the mutation rate was changed from 5% to 6%. The IGA evolved designs over ten iterations. At each iteration, sixteen mobile phones or blood glucose meters were shown (depending on the trial). In the first nine iterations, participants selected their four favorite designs, in the last iteration, they selected their single favorite. Figure 4.3 and Figure 4.4 show the user interfaces with blood glucose meters and mobile phones.

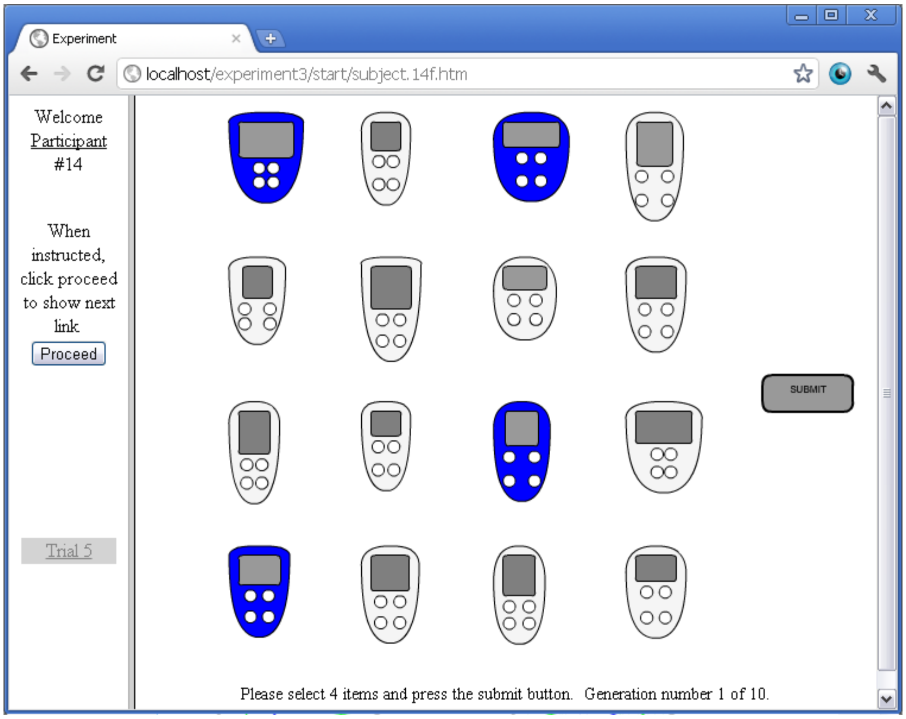


Figure 4.3. Screen capture of IGA interface with four blood glucose meters selected, denoted by a darker background.

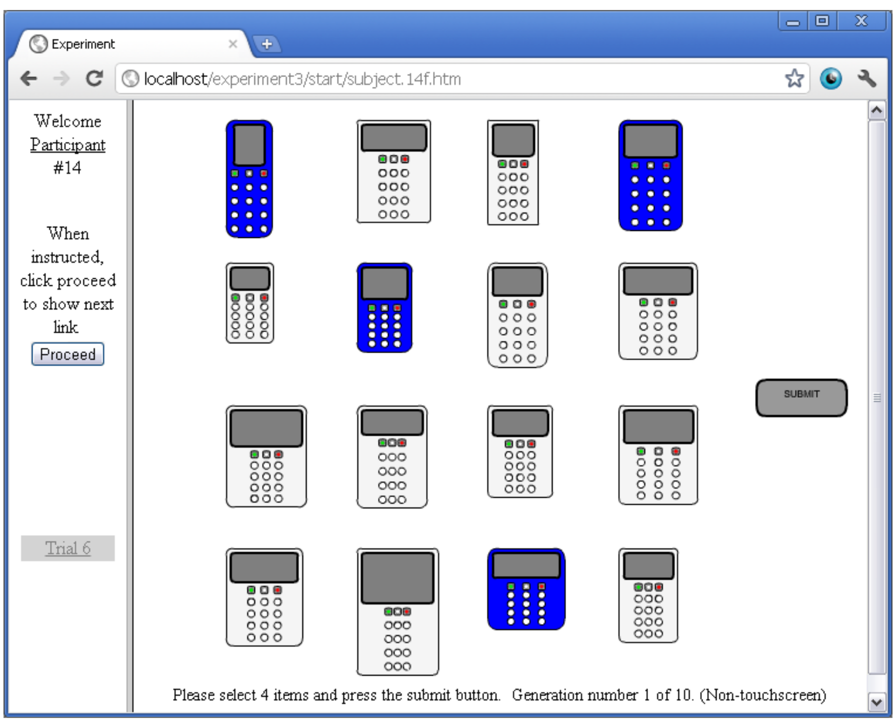


Figure 4.4. Screen capture of IGA interface with four mobile phones selected, denoted by a darker background.

The IGA was coded in PHP, a popular scripting language, and read into Adobe Flash to render the designs as a webpage. PHP recorded the displayed values and user feedback and saved it to MYSQL databases, an online database system. Variable ranges were set in PHP. The user interface webpage, Figure 4.3, displayed the participant's number, the current trial, and a link to proceed to the next trial or survey. A Flash webpage instructed users with a prompt of their goal before each trial.

Procedure

Participants were directed to the experiment website via a recruitment email to university department list-serves. Interested participants were qualified via a screening form on the experimental website. Qualified participants were given instructions and a link via email to the study website where they filled out a consent form.

After consent, participants were given an overview of the study and instructed on how to access the study online and use the software tool. Participants were given a small overview of what blood glucose meters are, and told the most important thing for them to do was to pick designs based ONLY on aesthetics the prompt for that trial, and not based on previous prompts.

Participants completed eight trials of the IGA software; two practice (one mobile phone and one blood glucose meter), and then six test trials (alternating between blood glucose meter, and mobile phone design). After each trial was a short questionnaire, and after all eight trials was a longer debrief questionnaire. Participants completed the study outside of the lab on a computer meeting the inclusion requirements.

Data analysis

After participants completed the study, results were pulled from the MYSQL database and saved locally to Excel (Microsoft Corporation). Statistical tests were performed in R (The R Project for Statistical Computing).

The normality of the distributions was tested with probability plots and Shapiro-Wilks W-Test. After determining that a portion of the results were non-parametric, Wilcoxon Signed Rank Test was used to test within-subject variables and the Wilcoxon Rank Sum Test (equivalent to the Mann-Whitney U test) was used to test the location variable. Principal Component Analysis (PCA) was performed to determine which variables contributed to the variance between participants.

Results

Twelve male and ten female participated in Experiment One. Eighteen of the participants had some or extensive blood glucose meter experience, four did not. Participants consistently said that their final selection was one of the best for that trial (94%, n=130).

A Dunnett's test without assuming equal variance was performed in R testing means between the mobile phone and blood glucose meter designs. All of the pairwise comparisons found the groups to be statistically significantly different (corrected to an $\alpha \leq 0.05$) on all variables except: screen width (p-value .1028), design radius (p-value .5782), the ratio of screen area to button area (p-value .3931), the overall height of the design (p-value .9210), and the width:height ratio of the screen (p-value .9934), which were all found to be similar across mobile phones and blood glucose meters.

PCA tests of the independent variables were not conclusive with the first components representing 31% and 41% of variance for mobile phones and blood glucose meters respectively, while the second components represented 22% and 21% respectively. Both PCAs found the first components' variance coming from a balanced mix of all of the independent variables, except for device radius, and the second component dominated by the radius variable. PCAs did not show a clear measure of beauty on which participants varied either. A PCA of the ratios and weightings identified as classical measures of beauty did not show a clear measure driving variance among participants.

Participants selected a wide range of mobile phone and blood glucose meter designs. Figure 4.5 and Figure 4.6 show the smallest, mean, and largest designs of the selected designs on the last generation for mobile phones and blood glucose meters. Some participants did not complete all of the studies correctly, so the mean values shown are a mean of each participant's mean for blood glucose meters or mobile phones respectively. Several classic definitions of beauty were compared between the designs; mean and standard deviations are shown in Table 4.2.

Table 4.2. Mean (SD) of aesthetic values from Experiment One.

Measure	Phone	Meter
Screen area : button area ratio*	1.512 *(0.488)*	1.369 *(0.343)*
Screen width : screen height ratio*	1.171 *(0.281)*	1.200 *(0.256)*
Screen Width*	36.228 *(7.180)*	33.140 *(6.137)*
Design Radius*	7.816 *(1.855)*	8.531 *(2.997)*
Overall Height of design*	76.877 *(6.916)*	78.056 *(6.282)*
Height : width ratio	1.600 *(0.188)*	1.766 *(0.246)*
Keypad width : keypad height ratio	0.714 *(0.046)*	1.028 *(0.095)*
Screen height : keypad height ratio	0.968 *(0.191)*	1.087 *(0.165)*
Screen width : keypad width ratio	1.543 *(0.309)*	1.250 *(0.205)*
Button horizontal spacing: vertical spacing ratio	2.086 *(0.962)*	1.273 *(0.673)*

** Indicates that this variable was similar between designs (not significantly different at a corrected α of 0.05)*

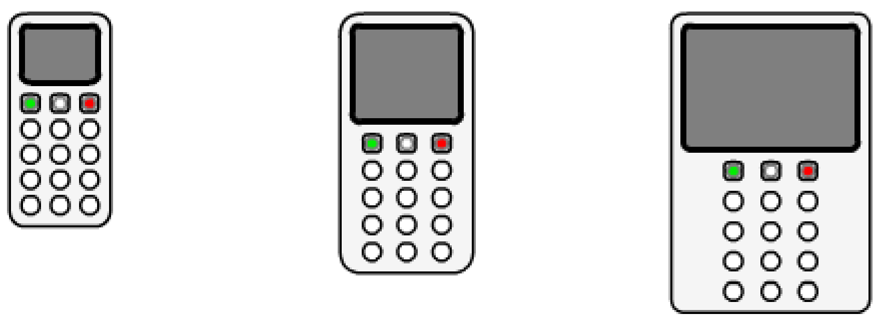


Figure 4.5. Select mobile phone designs, from left to right; the smallest final generation design selected, the mean design across all participants, and the largest design across all participants.

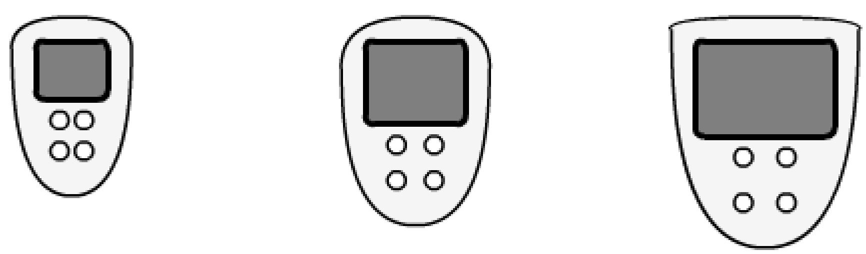


Figure 4.6. Select glucose meter designs, from left to right; the smallest final generation design selected, the mean design across all participants, and the largest design across all participants.

Discussion

Using a study tool validated in previous reduces the need to verify that the IGA converges and exceeds discriminability. Subjective data reaffirmed that this IGA is non-deterministic, exceeds user discriminatability, and converges towards a solution.

Dunnett's comparison of means without assuming equal variance between the design prompts easily provided corrected *p*-values without being overly conservative. More strict corrections, like Bonferroni, may have unfairly pushed more design options to be not significantly different. Dunnett's tests were also significantly more straightforward to perform and analyze than compiling a series of paired t-tests.

Comparing the mobile phone and blood glucose meter design prompts, we see that several variables and aesthetics ratios are not significantly different between the designs. The values of screen width, design radius, and overall height were not found to be significantly different, but are hard to interpret across designs because they may relate to an issue of scale between devices. The width-to-height ratio of the designs and the screen ratio between the screen area and the button keypad area was also not significantly different. Noting these two variables are not dissimilar may be useful for designers in deciding how to design new home healthcare devices.

Principal Component Analysis has been used to determine where participants diverged in preference (Nathan-Roberts & Liu, 2010), but the PCA results found here were not conclusive. Perhaps a PCA of smaller subgroups, such as only non-touchscreen mobile phone owners may provide a subset among which a PCA could find a smaller subset of variables along which participants varied.

Aesthetic preference of new products is critical to product success and a difficult area to get right (Veryzer, 1998). Using a similar product may reduce the inherent design risk in a new product's design by building on user's current aesthetic in another product group. Comparing user mobile phone and blood glucose meter aesthetic preference in this study did not provide a simple model which can be used to compare device aesthetics across these domains. Further testing of more similar and dissimilar designs as well as novice and experienced users will further help shed light on the connections between these devices.

Chapter Four, Experiment Two

Experiment One tested how closely tied user design preferences are for devices used in a similar manner and environment, as well as the effect of experience on that connection. Experiment Two seeks to further test both of these relationships on a wider array of designs and a larger novice participant set. The goals of this study are to further understand how closely aesthetic preference can be connected across devices for novice and experienced user groups.

Method

Experiment Two further tests the connection between the two device types by adding two more blood glucose meter designs to the mix, one more similar to a mobile phone, and one ostensibly less similar. The software used in Experiment Two evolved designs in the same manner as explained in Experiment One.

Participants

Participants were recruited from the college community, 35 participants; 16 males and 19 females participated in this study. The mean age was 22 years, with a standard deviation of 4.3 years. Roughly half (18) of the participants had experience using a blood glucose meter, 17 did not. The inclusion criteria were the same as Experiment One. The most popular phone amongst participants was the Apple iPhone 4 (eight participants), followed by the Apple iPhone 3 (four participants). Of the participants who did not have Apple iPhones, three used an Apple iPod (model not specified) regularly (defined as more often than “several times per month”).

Variables

Within-subjects, Experiment Two added two additional IGA conditions; blood glucose meters that were more similar to the phones displayed than the previous experiment’s designs, and ones that were more dissimilar to the phones displayed. The IGA in Experiment Two used the same independent variables as the previous study, namely; horizontal screen size, vertical screen size, horizontal button spacing, vertical button spacing, and the radius of the device and the screen. The independent variables are shown on the designs in Figure 4.7. The dissimilar case, blood glucose meter style A, was rotated, had directional or colored buttons, had a thinner screen outline, and a strip port added to the right side. The strip port, where test strips are inserted, are a common component of the front fascia of a blood glucose meter. In the design for this study, the strip port scaled with the radius.

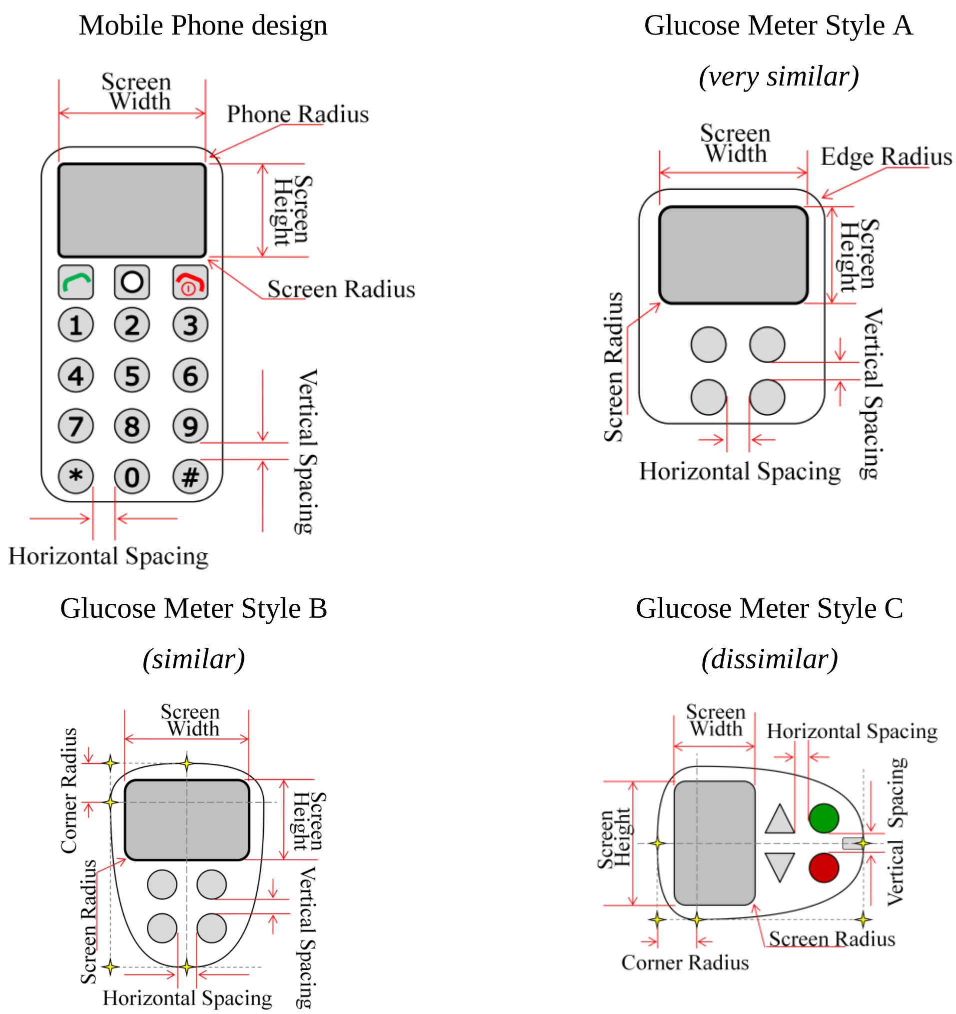


Figure 4.7. The four designs used in Experiment Two. Each uses the same independent variables. Blood glucose meter styles B and C use quadratic Bézier curves between the corners and points highlighted with stars to draw the shell around the screen and buttons.

Interactive Genetic Algorithm Configuration

The IGA used here was configured in the same manner as explained in Experiment Two, Chapter Three on page 28. The mutation rate was 6%, the same as the first experiment in this chapter.

Procedure

Participants were directed to the experiment website via a recruitment email to university department list-serves. Interested participants were qualified via a screening form on the experiment website. Once qualified, participants received information, filled out a consent form, received training, and completed the study remotely.

After consent, participants were given instructions about the tool, blood glucose meters, and how to complete the study. Participants were instructed on the importance of making "their choices based ONLY on aesthetics, and to consider ONLY the current prompt." After training, participants completed a series of trials with a short questionnaire after each trial, and a longer debriefing survey after the last trial.

Participants completed twenty trials; two practice of each design (Phone, Blood Glucose Meter Style A, Blood Glucose Meter Style B, and Blood Glucose Meter Style C), and three test trials of each design. Participants first completed the practice trials in the order mentioned above, but test trial order was randomized. After all of the participants had completed the study, the data were analyzed in the same manner as in Experiment One. The prompts used are shown in Table 3.5.

Table 4.3. Prompts used for Experiment Two in IGA software.

Goal	Prompt
Practice	"The software tries to learn from your previous selections. Please experiment with the software by picking designs that you like and see how the program responds using your selection to give you new options in the next step."
Phone trial	"This trial will be a phone trial. On every iteration please select the design(s) based ONLY on what you find MOST aesthetically pleasing."
Style A blood Glucose Meter	"This trial will be a glucose meter trial (Style A). On every iteration please select the design(s) based ONLY on what you find MOST aesthetically pleasing."
Style B blood Glucose Meter	"This trial will be a glucose meter trial (Style B). On every iteration please select the design(s) based ONLY on what you find MOST aesthetically pleasing."
Style C blood Glucose Meter	"This trial will be a glucose meter trial (Style C). On every iteration please select the design(s) based ONLY on what you find MOST aesthetically pleasing."

Results

Participants were split by glucose meter use experience; 18 had blood glucose meter usage experience on themselves or a loved one, and17 did not have any blood glucose meter experience. A paired t-t test of each prompt type comparing means of the experienced and inexperienced blood glucose meter users (i.e., meter style C experienced users compared against inexperienced users) found a number of statistically significant differences between the groups (Table 4.4). All of these variables were no longer significantly different after being corrected for multiple comparisons using an approximate multivariate *t*-distribution with the Satterthwaite approximation to determine the degrees of freedom.

Table 4.4. Uncorrected paired t-test results comparing groups by glucose meter experience. All other were not significantly different (at an α equal to 0.05), and all variables here had *p*-values above 0.05 after being corrected for multiple comparisons.

Prompt type	Variable	Uncorrected *p*-value
Phone	Button spacing (width)	0.0157
Phone	Radius	0.0092
Phone	Keypad area	0.0115
Phone	Keypad width	0.0157
Meter A (similar)	Screen height	0.0425
Meter A (similar)	Screen width	0.0452
Meter A (similar)	Screen area : keypad area ratio	0.0352
Meter A (similar)	Overall width *(driven by the greater of screen size or horizontal button spacing)*	0.0452
Meter B (mid)	Screen height	0.0417
Meter B (mid)	Overall height	0.0330
Meter C (dissimilar)	Screen width	0.0039
Meter C (dissimilar)	Screen area	0.0115
Meter C (dissimilar)	Screen area : keypad area ratio	0.0498
Meter C (dissimilar)	Overall height	0.0035
Meter C (dissimilar)	Screen width : keypad width ratio	0.0197
Meter C (dissimilar)	Screen height : screen width ratio	0.0387

The ability of this IGA to converge towards a desired solution and exceed participant discriminability has been quantitatively tested elsewhere, qualitatively, 94.0% said that their “final selection among the best presented in this trial given [their] goal(s),” and 83.6% said that the designs start[ed] looking the same” (n=422).

Comparing the mobile phone design in one-to-many Dunnett’s t-tests of the independent and resultant aesthetic variables means against the three meter types found the groups to be statistically significantly different (corrected to $\alpha \leq 0.05$) for all variables except those shown in Table 4.5.

Table 4.5. Dunnett's pairwise comparisons of the difference between the mobile phone prompt and each type of blood glucose meter design corrected for multiple comparisons. A higher *p*-value indicates a stronger correlation between mobile phone values and blood glucose meter values for that variable. (Only tests with a *p*-value α above 0.05 are included here.)

Design compared to phone	Variable	Corrected *p*-value
Meter A (similar)	Width	$0.9972^{A, B, C}$
Meter A (similar)	Screen height : keypad height ratio	$0.8228^{A, B, C}$
Meter A (similar)	Screen width : keypad width ratio	$1.0000^{A, B, C}$
Meter A (similar)	Screen width : screen height ratio	$1.0000^{A, B, C, *}$
Meter B (mid)	Width	$0.9993^{A, B, C}$
Meter B (mid)	Screen height : keypad height ratio	$1.0000^{A, B, C}$
Meter B (mid)	Screen width : keypad width ratio	$0.9403^{A, B, C}$
Meter B (mid)	Screen width : screen height ratio	$1.0000^{A, B, C, *}$
Meter B (mid)	Radius	$1.0000^{B, C, *}$
Meter B (mid)	Keypad width : keypad height ratio	0.1049
Meter C (dissimilar)	Width	$0.6351^{A, B, C}$
Meter C (dissimilar)	Screen height : keypad height ratio	$1.0000^{A, B, C}$
Meter C (dissimilar)	Screen width : screen height ratio	$1.0000^{A, B, C, *}$
Meter C (dissimilar)	Screen width : keypad width ratio	$0.4332^{A, B, C}$
Meter C (dissimilar)	Radius	$0.9760^{B, C, *}$
Meter C (dissimilar)	Height	0.9997^{*}
Meter C (dissimilar)	Overall height : width ratio	0.4681

Superscript letters highlight variables that are shared across meter designs

* Indicates pairwise comparison was also not significantly different in Experiment One

PCA tests of the independent variables were not conclusive with the first components representing 38%, 29%, 35% and 29% of variance for mobile phones, blood glucose design meter A, design B, and design C respectively, while the second components represented 22%, 23%, 22% and 24% for the same designs respectively. PCAs did not show a clear measure of beauty as the locus of where participants varied either. A PCA of the ratios and weightings identified as classical measures of beauty did not show a clear measure driving variance among participants.

Participants selected a wide range of mobile phone and blood glucose meter designs. Figure 4.8 and Figure 4.9 show the smallest, mean, and largest designs of the selected designs on the last generation for mobile phones and blood glucose meters.

Some participants did not complete all of the studies correctly, so the mean values shown are a mean of each participant's mean for blood glucose meters or mobile phones respectively. Several classic definitions of beauty were compared between the designs; mean and standard deviations are shown in Table 4.6.

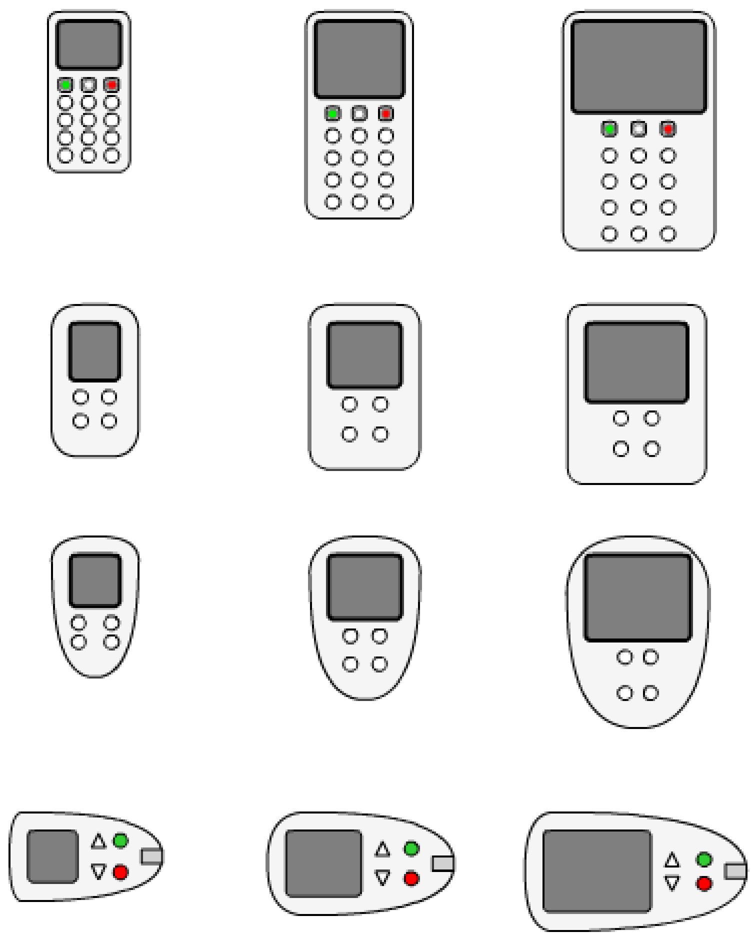


Figure 4.8. Select designs, from left to right; the smallest selected final generation design, the mean design across all participants, and the largest selected design from all participants.

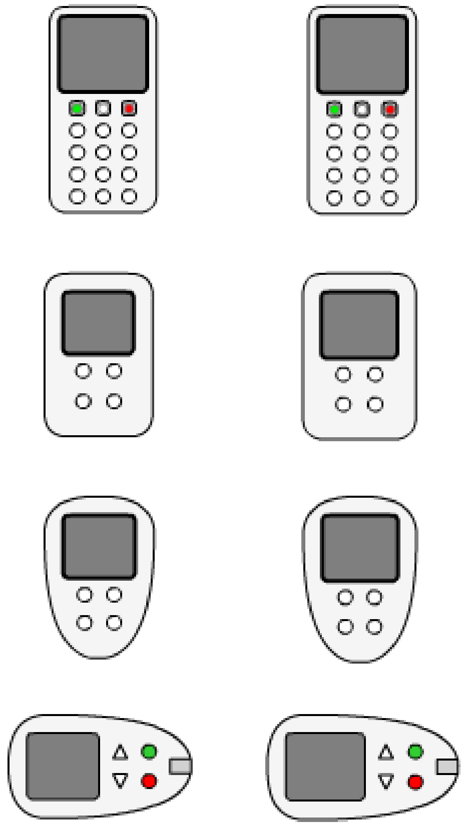


Figure 4.9. Comparison of glucose meter experience effect, which was not significant after correction. Mean final selected designs; left, designs from participants without blood glucose meter experience, right, designs from participants with blood glucose meter experience.

Table 4.6. Range of aesthetic values from Experiment Two, value (SD).

Variable	Phone	Meter A	Meter B	Meter C
Screen : button area ratio	1.626 (0.538)	2.431 (0.814)	2.604 (0.812)	2.688 (0.822)
Height : width ratio	1.740 (0.241)	1.515 (0.194)	1.500 (0.206)	1.822 (0.251)
Keypad width : keypad height ratio	1.094 (0.082)	1.024 (0.135)	1.034 (0.174)	1.017 (0.136)
Screen height : keypad height ratio	1.541 (0.268)	1.469 (0.286)	1.519 (0.248)	1.538 (0.292)
Screen width : keypad width ratio	1.620 (0.346)	1.633 (0.342)	1.709 (0.446)	1.751 (0.423)
Button horizontal spacing: button vertical spacing ratio	1.732 (0.719)	1.112 (0.406)	1.232 (0.692)	1.071 (0.476)
Screen width : screen height ratio	1.162 (0.227)	1.152 (0.260)	1.153 (0.266)	1.176 (0.325)

Discussion

In terms of total trials ran, study two was the largest study to date using this web-based IGA software. The IGA functioned well, and participants found it converged and exceeded their ability to discriminate between designs. While there were significantly different values between the users with blood glucose meter use experience, and those without, they were not found to be significant after correction for multiple comparisons. This difference is likely an artifact of the sample size. Had more subjects participated in the study, it is possible that there would still be a difference, but the difference between phone designs would be smaller, or nonexistent (phone ownership and blood glucose meter experience was not compared). The sheer number of variables that were significantly different based on experience before correction supports the hypothesis that experience has an effect on design preference.

Similar to Experiment One, Dunnett's comparison of means were used to highlight key similarities between the design prompts. All of the blood glucose meter designs were similar to the phones for ratios that included screen height and width and keypad height and width, as well as just screen height or just the overall width of the designs. The statistical similarity of the overall width of the mobile phone and blood glucose meters is likely an artifact of the way they were designed, resulting in similar ranges of widths. The ratios which were not different between the mobile phone prompt and the blood glucose meter prompts are more interesting because they give insight into how the designs are similar independent of their actual size. It is clear from the strong similarities found across blood glucose meter designs that the proportions of the screen are very important to the overall design. Interestingly, the design prompt most similar to the phone, Meter A, also had the fewest variables that were similar to the mobile phone design. The Meter C, or dissimilar design, had the most variables that were statistically similar to the mobile phone design, whereas Meter A, or the similar design, had the fewest. Radius similarities between blood glucose meter designs B & C and mobile phone designs, like width, seem like artifacts of the geometry built into the IGA to model the devices. Similarly, the height of meter design C, which is defined as the horizontal height (because it is rotated), is also likely an artifact of the sizes of the components and geometry used to make a dissimilar design which used quadratic Bézier curves to form its exterior shape.

The ratios of dimensions can, and should be used for comparison with known aesthetics ratios mentioned earlier. The smallest standard deviation was found in the keypad width : height ratio, followed by the other two height width ratios, the overall height : width ratio and screen width : height ratio respectively. The keypad width: height ratio is very close to 1:1 (symmetry), even on the mobile phone designs, which have a non-symmetrical button layout; three across, and four high, opposed to the blood glucose meters which were an even two by two. Also consistent across design prompts were the ratios of heights & widths (screen height : keypad height, and screen width : keypad width), which were both close to the golden rectangle (1.618), and the square root of three (1.732). The ratios most dissimilar between the phone design prompt and the blood glucose meter prompts are the ratio of screen area to keypad area and the horizontal and vertical button spacing ratio. The majority of the aesthetics ratios found in Experiment Two were similar to those found in Experiment One. Of the ones that were dissimilar, they seemed to stem from Experiment Two having a shorter average keypad height, making for different screen height : keypad height ratios for both devices, a different keypad width: keypad height ratios for phone designs, and a different screen :button area ratio for blood glucose meters. Some research has debated the notion that the golden section, or golden ratio is value of universal beauty, or that it enjoys a place of privilege (Sudweeks, 1999), but these results confirm that while the exact value may not be correct, and as posited elsewhere, ratios may be driven by context (Vanschaik & Ling, 2003), a ratio similar to the golden section, or square root of three, helps predict these designs.

Principal Component Analysis failed again to point to a single variable or small group of variables as the primary axis along which participants vary in preference. Instead, the participants varied along all of the variables, with none being the driving factor in the variance between designs. While it makes the results harder to interpret for a designer, these results highlight the importance of each of the variables chosen.

The second experiment shed further light on the extent to which users have similar aesthetic preferences for two similar devices, independent of their experience with the device. The similarity of the aesthetic ratios and comparison of means shows that while not universally the same, using layout ratios for similarly used devices, even from other domains, can provide a good starting place for designs.

General Discussion

Combined, these studies point to users having an aesthetic link between devices across domains. On a macro-scale there are clear similarities between these two devices which have very different use goals. The link is not unequivocal, and is stronger for several variables or ratios than others. Specifically the shape of screen and how its dimensions related to keypad dimensions were found to be connected across domains in both studies. Also important to note is the similarity of many of the ratios in both studies.

While there are unmistakable connections between the mobile phone and blood glucose device designs, the connections are not absolute. There are clear differences in their actual sizes and their layouts for obvious, functional reasons as well as because of the geometry preset by the experimenters within the IGA for each device prompt. The more minute values, like the individual button spacing, or symmetry between horizontal and vertical button spacing were not as closely tied between devices as the more macro values such as the keypad height : width ratio.

The relative fidelity of the Experiment Two, “Meter C” or dissimilar prompt, which had shaped and colored buttons as well was showing the port for the glucometer test strips may have helped participants envision the final design better than the other blood glucose meter designs (Meter A, and Meter B). This slight increase in fidelity, like engineering constraints on a real world design, likely helped participants design blood glucose meters that were more realistic. Although user preference for one meter prompt over another was not recorded, it is possible that the fidelity would have an interesting interaction with attractiveness, as found in Sauer & Sonderegger, 2009, where less attractive designs that were lower fidelity were rated as more attractive than their higher fidelity genesis, but the effect was not present for more attractive designs. It is also likely that participants would reject the similar blood glucose meter design (Meter A) because of its simplicity, striving more for the Gestalt idea of order-within-variety, or enough complexity to be interesting, as has been found for art (McWhinnie, 1987).

At the heart of these experiments is desire to address the question of a new product development. New product developments are tricky business, especially discontinuous new product development. In the testing of new discontinuous products designers find customers resistant to the new products, wary of their safety, or focusing on what designers think are unimportant features (Veryzer, 1998). With all of these difficulties, significant effort has been invested in designing products quantitatively based on similar products (Smyth & Wallace, 2000). The methods outlined in this chapter can be used to by designers to quickly get structured, quantitative input from a large set of users. When using an IGA, or any other user-driven design tool to design a new, radical device, it is crucial that designers sufficiently educate users on the function of the device, and how it is used. Users are able to hold multiple goals in mind and generate designs which are multi-objective optimizations (Chapter 3), it is therefore incumbent upon designers to make sure that the users fully grasp the new product's purpose and method of use.

Separate from using IGAs to design radically new devices, it is also possible to use IGAs to perform mass customization. Advances in the rapid prototyping, and Computer Numerical Control (CNC) machining in the last decade allow for individuals to make custom multi-faceted designs more quickly, easily, and cheaper than ever before.

However, before these results are used to improve or design the next widget, it is important to heed several cautions. First, the designs shown here are snapshots of the current trends and state of the art. A comparison between conventional, corded phones and mobile phones in the early nineteen nineties may have revealed significantly larger phone designs, or different aesthetic preferences. This chapter does not test the temporal affect of these aesthetic ratios. Second, it would be wrong for a designer to use the exact values or even the same independent variables, without undergoing several steps to determine the variables (Liu, 2003; Nathan-Roberts, et al., 2009; Nathan-Roberts & Liu, 2010). Third, experience with a technology can, in some cases have an effect on their design preference (Nathan-Roberts, Kelly, & Liu, 2011), so it is important to recruit the same participants that would likely be adopters of the future design, and have a wide range of experiences.

Visual appearance is not just a hedonic issue, but visually appealing designs also have a positive effect on performance (Sonderegger & Sauer, 2010), and on perceived difficulty (Pomales-Garcia, Liu, & Mendez, 2005). This study represents an important step into quantitative methodologies of linking aesthetics across domains, but the further testing of the link between devices across domains and the effect of experience are important to further aiding the design of radically new products. Other areas of future work include measuring if users are more likely to use devices they designed themselves using methods like the Technology Acceptance Model (Szajna, 1996), adding additional functionality to the IGA tool to allow it to take physical ergonomics into account as well.

References

Bauerly, M., & Liu, Y. (2009). Evaluation and Improvement of Interface Aesthetics with an Interactive Genetic Algorithm (IGA). *International Journal of Human Computer Interaction* , 25 (2), 155-166.

Centers for Disease Control and Prevention. (2011*). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States*, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Cons, N., & Jenny, J. (1994). Comparing perceptions of dental aesthetics in the USA with those in eleven ethnic groups. *Int Dent J.* , 44 (6), 489-94.

Gero, J. S., & Jun, H. J. (2006). *Visual Semantics Emergence to Support Creative Designing: a Computational View*. Australia: University of Sydney.

Gong, J., & Tarasewich, P. (2004). Guidelines for handheld mobile device interface design. *2004 DSI Annual Meeting* (pp. 3751-3756). In Proceedings of the 2004 DSI Annual Meeting.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S. W., & Kim, J. (2004). Identifying Mobile Phone Design Features Critical to User Satisfaction. *Human Factors and Ergonomics in Manufacturing* , 14, 15-29.

Harrington, S. J., Naveda, J. F., Jones, R. P., Roetling, P., & Thakkar, N. (2004). Aesthetic Measures for Automated Document Layout. *ACM Symposium on Document Engineering* (pp. 109-111). Milwaukee, Wisconsin, USA: ACM.

Hughes, M. D. (2009). The Business of Self-Monitoring of Blood Glucose: A Market Profile. *J Diabetes Sci Technol* , 3 (5), 1219-1223.

ITU. (2010). *Measuring the Information Society*. International Telecommunication Union. Geneva, Switzerland: ITU.

Karlsson, P., & Djabri, F. (2001). Analogue Styled User Interfaces: An Exemplified Set of Principles Intended to Improve Aesthetic Qualities in Use. *Mobile HCI 2001: Third International Workshop on Human-Computer Interaction with Mobile Devices.*

Kelly, J. C., Papalambros, P., & Wakefield, G. (2006). The Development of a Tool for the Preference Assessment of the Visual Aesthetics of an Object Using Interactive Genetic Algorithms. *9th Generative Art Conference*. Milan, Italy.

Licitra, E. (2011, April 29). *Course ID 2012: Glucometer*. Retrieved June 23, 2011, from Georgia Tech College of Architecture: http://www.id.gatech.edu/node/145686

Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. Ergonomics , 46, 1273-1292.

Lloyd, L. (2010, April). Survival of the Smartest. *MD+DI; Medical Device and Diagnostic Industry* , 32 (4).

Marcus, A. (1992). *Graphic Design for Electronic Documents and User Interfaces.* New York: ACM Press.

Mastrangelo, L. (2011, May 1). *Course ID 2012: Sentinel Blood Glucose Monitor - Form and Function Project*. Retrieved June 23, 2011, from Georgia Tech College of Architecture: http://www.id.gatech.edu/node/146492

McWhinnie, H. J. (1987). Some Studies of Aesthetic Preference. *The British Journal of Aesthetics* , 76-86.

Nathan-Roberts, D., & Liu, Y. (2010). Investigation of Relative Mobile Phone Size Preference Using Interactive Genetic Algorithms. *54th Annual Meeting of the Human Factors and Ergonomics Society.* San Francisco, California: HFES.

Nathan-Roberts, D., & Liu, Y. (2012). Testing Aesthetic and Function Design Preference for Touchscreen and Non-Touchscreen Mobile Phones Using Interactive Genetic Algorithms. *In review* .

Nathan-Roberts, D., Beeker, A., & Liu, Y. (2009). Modeling Two Key Physical Ergonomic Problems with Mobile Phones. *University of Michigan Engineering Graduate Symposium*. Ann arbor.

Nathan-Roberts, D., Kelly, J. C., & Liu, Y. (2011). Determining the Effect of Users' Mobile Phone on Design Preference via Interactive Genetic Algorithms. *55th Annual Meeting of the Human Factors and Ergonomics Society*. Las Vegas, Nevada: HFES.

Nathan-Roberts, D., Liu, Y., Young, M., Hays, J. D., An So, C., Little, C., et al. (2009). Content Analysis of Literature on Mobile Phone Design Features. *17th World Congress on Ergonomics.* Beijing, China: IEA.

Ngo, D. C., & Byrne, J. G. (2001). Another Look at a Model for Evaluating Interface Aesthetics. *Int. J. Appl. Math. Comput. Sci. , 11* (2), 515-535.

Noblet, J. (1993). *Industrial Design: Reflections of a Century*. Paris: Abbeville Press, Inc.

Obrador, P. (2007). Content Selection Based on Compositional Image Quality. *IS&T/SPIE 19th Annual Symposium on Electronic Imaging*. San Jose, California: SPIE.

Otto, K. N., & Wood, K. L. (2001). *Product Design: Techniques in Reverse Engineering and New Product Development*. Prentice Hall.

Plos, O., & Buisine, S. (2006). Universal design for mobile phones: a case study. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 1229-34). Montréal, Québec, Canada: ACM.

Pomales-Garcia, C., Liu, Y., & Mendez, D. (2005). Web-Based Distance Learning Technology: Does Appearance Matter. *Human Factors and Ergonomics Society Annual Meeting Proceedings* , 49, 755--758.

Rogers, W. A., Mykityshyn, A. L., Campbell, R. H., & Fisk, A. D. (2001, Winter). Analysis of a Simple Medical Device. *Ergonomics in Design* , 9 (1), pp. 6-14.

Sauer, J., & Sonderegger, A. (2009). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics* , 670-677.

Seva, R. R., Duh, H., & Helander, M. G. (2006). Factors Affecting Emotional Responses in Mobile Phone Design Evaluation. *Human Factors and Ergonomics Society 50th Annual Meeting*, (pp. 2173-2177). San Francisco.

Smyth, S. N., & Wallace, D. R. (2000). Towards the Synthesis of Aesthetic Product Form.

Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* , 41, 403--410.

Story, M. F., Luce, A. C., & Rempel, D. M. (2009). Accessibility and Usability of Diabetes Management Devices for Users with Vision Disabilities. *Human Factors and Ergonomics Society Annual Meeting Proceedings* , 53 (11), 699-703(5).

Sudweeks, F. &. (1999). Quantifying Beauty: An Information System for Evaluating Universal Aesthetics. *2nd Western Australian Workshop on Information Systems Research (WAWISR).* Perth, W.A.

Szajna, B. (1996). Empirical Evaluation of the Revised Technology Acceptance Model. *Management Science* , 42, pp. 85--92.

Vanschaik, P., & Ling, J. (2003). The Effects of Screen Ratio and Order on Information Retrieval in Web Pages. *Displays* , 24 (4-5), 187-195.

Veryzer, R. W. (1998). Key Factors Affecting Customer Evaluation of Discontinuous New Products. *Journal of Product Innovation Managemen*t , 15 (2), 136-150.

Wild, S., Roglic, G., Green, A., Sicree, R., & King, H. (2004). Global Prevalence of Diabetes; Estimates for the year 2000 and projections for 2030. *Diabetes Care* , 27 (5), 1047-1053.

Chapter 5

Integrating Aesthetic and Usability Factors in the Design of Mobile Phones and Blood Glucose Meters

Introduction

Studies have shown that aesthetics has a large impact on mobile phone selection (Han, Kim, Yun, Hong, & Kim, 2004), specifically that the size and shape of the device and the individual components have the largest affect on mobile phone aesthetics (Chapter 2). Preference of size and shape differs between people along several key dimensions. Prior work indicates that changes in horizontal and vertical screen dimensions account for a large portion of the variability between people's aesthetic preference (Chapters 2, 3, 4).

Goals

The goals of these studies are to test the ability of an aesthetics IGA and physical ergonomics GA to work together to optimize a design based on multiple constraints. These studies also seek to enrich the sometimes separate fields of aesthetics and physical ergonomics. By providing a simple model of physical ergonomics for designers they can balance their design sense with safety.

Chapter Five, Experiment One

Method

Experiment One tests the integration of a Genetic Algorithm (GA) with the previously tested Interactive Genetic Algorithm (IGA). The goals of the experiment are to validate the ability to integrate a GA with an IGA, and to test the effect of designing a mobile phone for aesthetics and physical ergonomics.

Participants

All of the participants, except for one were engineering students at the University of Michigan. Of the 32 students, 17 were male, and 15 were female. The mean age was

20 years, with a standard deviation of 1.2 years. Participants used a wide variety of mobile phones; the most common was the Apple iPhone 4 (12 participants). None had been a participant in prior IGA studies. The inclusion criteria were ownership or extensive use of a mobile phone, access to a computer with a high-speed internet connection and 19-inch to 21-inch monitor, and a lack of disabilities that would prevent them from safely using a computer for the duration of the study.

Variables

The independent variables are similar to those used in the exploratory study explained in Chapter 2, with the addition of the within-subject variable of algorithm-type; Interactive Genetic Algorithm combined with a Genetic Algorithm, or solely an Interactive Genetic Algorithm.

The within-subject independent variables that were used to generate a computer rendered mobile phone previously were used here again. The variables were the horizontal and vertical screen dimensions, the horizontal and vertical button spacing, the roundness (the radius of the outside of the mobile phone or blood glucose meter and radius of the screen). The radius variable varied the exterior radius and the screen radius simultaneously. The screen radius was one-half of the exterior radius value (Figure 5.1).

Within-subject, the design alternated between being designed with user selection as the only input to the IGA, or with the addition of a genetic algorithm fitness score of the usability of the designs included as well through a rudimentary ergonomics model.

At each trial participants were instructed to select the design(s) that were most aesthetically pleasing.

The most relevant control variables within the study were the minimum spacing between the screen and outer edge, the minimum spacing between the buttons and the outer edge, and the number of times participants experienced each prompt in the study (five).

The primary dependent variables of interest are the design and ergonomic rating of the selection in the last trial where the user could only select one design, and subjective participant preference measured through questionnaires at the end of each trial.

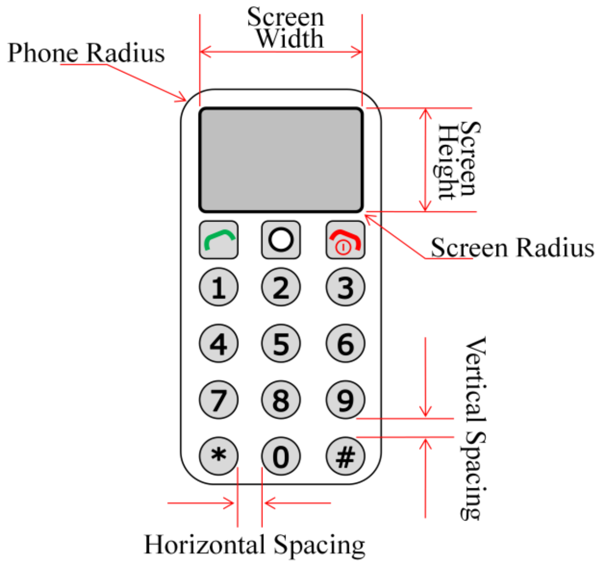


Figure 5.1. Independent dimension variables in mobile phone design

Algorithm Configuration

In the same manner as previous experiments, the algorithms in this study evolved designs over ten iterations. At each iteration, sixteen mobile phones were shown. In the first nine iterations, participants selected their four favorite designs, in the last iteration, they selected their single favorite. Figure 5.2 shows the web-based interface. The differences between this algorithm and previous algorithms in this dissertation was the inclusion of a Genetic Algorithm, which also necessitated a change to the roulette wheel allocation of selected designs, and a change mutation rate to ensure that the algorithms were not deterministic and converged.

The Genetic Algorithm (GA) was configured to create an ergonomic score for each design that was created by the IGA. The GA score used all of the independent design variables in the experiment to create a multi-objective equation that simultaneously optimized phone length, screen area, phone radius, and phone width for usability. The GA rated each objective on a score from one to ten, and the sum of the scores was considered the ergonomic score of the design (Equation 5.8). Phone length, screen area, and radius were considered "larger-the-better" design goals based on previous research; the GA awarded ten points to the maximum of the range, and decreased linearly to zero points at the minimum of the range. The phone width design goal was considered “middle-the-best” by the GA, giving ten points to the mean of the range, and reducing linearly to zero at either end of the range (Equation 5.7). Equation 5.1 to Equation 5.8 show how the GA performed its calculations.

$$DeviceLength_{\mathrm{i}} = \mathrm{ScreenHeight}_{\mathrm{i}} + \mathrm{VerticalButtonCount} \times (\mathrm{VerticalButtonSpacing}_{\mathrm{i}} - \mathrm{e})$$

Equation 5.1. Method for calculating device length for device i, i=1,…,16.. The vertical button count was driven by device type (five vertical buttons on the mobile phone design).

$$\mathrm{KeypadWidth}_{\mathrm{i}} = \mathrm{HorizontalButtonCount} \times (\mathrm{ButtonDiameter} + \mathrm{HorizontalButtonSpacing}_{\mathrm{i}}) - \mathrm{HorizontalButtonSpacing}_{\mathrm{i}}$$

Equation 5.2. Method for calculating keypad width for device i, i=1,…,16.. The horizontal button count was driven by device design (three horizontal buttons for the mobile phones designed).

$$\mathrm{DeviceWidth}_{\mathrm{i}} = \max(ScreenWidth_i, KeypadWidth_i)$$

Equation 5.3. Device Width calculation for device i, i=1,…,16.. Using the Screen Width variable and the calculated Keypad Width value for a device, the Device Width was equivalent to the larger of the two.

$$\boldsymbol{LengthScore_i} = 10 \times \left(\frac{DeviceLength_i - MinimumDeviceLength}{MaximumDeviceLangth - MinimumDeviceLength}\right)$$

Equation 5.4. Length Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Device Length and Maximum Device Length are the smallest and largest device lengths for the population respectively. This scoring method gives the longest device a score of 10, and the shortest a score of 0.

$$\boldsymbol{ScreenAreaScore_i} = 10 \times \left(\frac{(ScreenWidth_i \times ScreenHeight_i) - \mathrm{MinimumScreenArea}}{MaximumScreenArea - MinimumScreenArea}\right)$$

Equation 5.5. Screen Area Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Screen Area and Maximum Screen area are the smallest and largest products of the screen width and height for the population respectively. This scoring method gives the largest screen a score of 10, and the smallest screen a score of 0.

$$\boldsymbol{DeviceRadiusScore_i} = 10 \times \left(\frac{DeviceRadius_i - \text{MinimumDeviceRadius}}{MaximumDeviceRadius - MinimumDeviceRadius}\right)$$

Equation 5.6. Radius Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. Minimum Device Radius and Maximum Device Radius are the smallest and largest device radius values for the population respectively. This scoring method gives the largest radius a score of 10, and the smallest radius a score of 0.

$$\boldsymbol{DeviceWidthScore_i} = 10 \times \left(1 - \frac{\left|\text{DeviceWidth}_\text{i} - \text{MinimumWidth} - \frac{\text{MaximumWidth}-\text{MinimumWidth}}{2}\right|}{\frac{\text{MaximumWidth}-\text{MinimumWidth}}{2}}\right)$$

Equation 5.7. Device Width Score equation of the Genetic Algorithm fitness test for device i, i=1,…,16.. The Minimum Width and Maximum Width are the smallest and largest device width values for the population respectively. This scoring method gives devices with a width at the mean of widths a score of 10, and devices at either the maximum or minimum width a score of 0.

$$\boldsymbol{ErgonomicScore_i} = LengthScore_i + ScreenAreaScore_i + DeviceRadiusScore + DeviceRadiusScore$$

Equation 5.8. Overall Genetic Algorithm fitness function combining the other equations above.

The IGA would then take the Ergonomic Score into account in addition to the user selections for each IGA trial which included the GA. The IGA would include the top four GA scores as well as the top four user scores to determine the parents for the next generation. Each of the top four GA scores were given ten percent of a roulette wheel algorithm to determine the parents of each design for the next generation. Each of the four user selections were also given ten percent of the roulette wheel as well. The maximum percentage of the roulette wheel algorithm a single design could occupy was 20% (if it was one of the four most ergonomic designs and was selected by the user). The remaining designs split the remaining 20% of the roulette wheel algorithm. In the IGA only conditions each of the four user selected designs were awarded 20% of the roulette wheel algorithm, and the remaining designs split the remaining 20%.

Mutation was still applied to every variable as done in Chapter Four. Different mutation rates were tested before running the experiment to make sure that the algorithm explored the entire design space. Mutation rates from 5% to 20% were tested by the experimenter. A mutation rate of 12% was selected as a reasonably high mutation rate to ensure that the entire design space was explored, but still allowed the algorithm to converge.

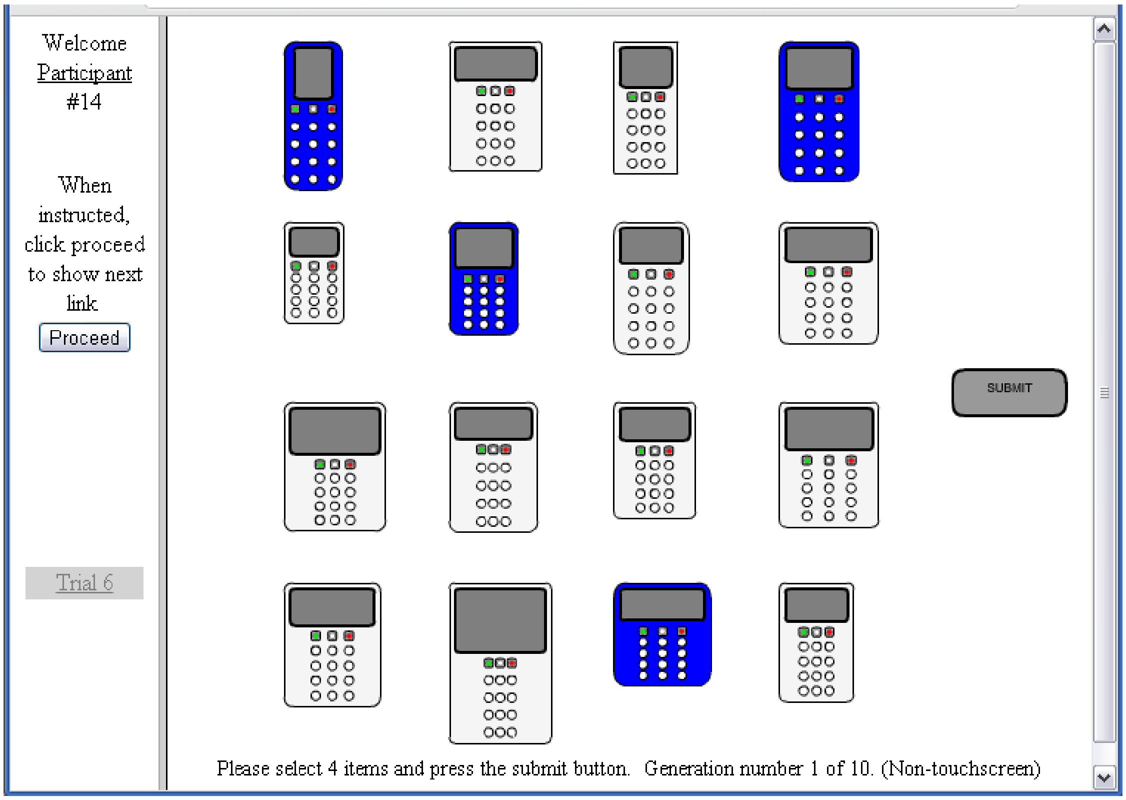


Figure 5.2. Screen capture of IGA interface with four mobile phones selected, denoted by a darker background.

The IGA and GA were coded in PHP, a popular scripting language, and read into Adobe Flash to render the designs as a webpage. PHP recorded the displayed values and user feedback and saved it to MYSQL databases, an online database system. Variable ranges were set in PHP. The user interface webpage, Figure 5.2, displayed the participant's number, the current trial, and a link to proceed to the next trial or survey. A Flash webpage instructed users with a prompt of their goal before each trial.

Procedure

Participants were directed to the experiment website via a recruitment email to university department list-serves. Interested participants were qualified via a screening form on the experimental website. Qualified participants were given instructions and a link via email to the study website where they filled out a consent form.

After consent, participants were given an overview of the study and instructed on how to access the study online and use the software tool.

Participants completed ten trials of the IGA software; two practice (one IGA with a GA and one IGA only), and then eight test trials (alternating between the two types of algorithms). After each trial was a short questionnaire, and after all ten trials was a longer debrief questionnaire. Participants completed the study outside of the lab on a computer meeting the inclusion requirements.

Results

Participant trials were split by the presence of a Genetic Algorithm as well as an Interactive Genetic Algorithm; half of the trials had only the IGA (n=79), while the other half had an IGA and GA functioning at the same time (n=79). Of the trials rated (n=152), 89.6% said that their "final selection among the best presented in this trial." Additionally, participants rated 82% (n=124) of the trials "started looking the same to you." On the trials where it did start looking the same, the mean was trial 5.8 of the 10 trials.

Comparing the first and last generations through paired t-tests, in Table 5.1, shows that they are statistically significantly different. Ergonomic score was not tested in this comparison as it was not relevant. Similarly, Table 5.1 shows that the designs created using an IGA alone are statistically significantly different from designs with a Genetic Algorithm as well for most variables, including ergonomic score.

Table 5.1. Paired t-test *p*-values comparing the first and last generations, and the difference between IGAs alone and IGAs with GAs as well.

	Button		Screen		Radius	Ergonomic
	Horizontal	Vertical	Vertical	Horizontal		Score
First to last generation*	5.061e-13	5.51e-7	2.2e-16	1.615e-5	2.2e-16	
IGAs compared to IGAs with a GA+	0.08531	1.29e-7	3.848e-5	0.1521	1.061e-7	6.904e-10

*$\alpha = 0.01$ after Bonferroni correction. +$\alpha = 0.00833$ after Bonferroni correction.

The presence of a Genetic Algorithm did not statistically significantly alter the aesthetic score of the designs (*p*-value 0.1367), as shown in Table 5.2. It did change the ergonomic score, *p*-value 6.904e-10, also shown in Table 5.2. The difference in mean values can be seen in Table 5.3.

Table 5.2. Mean aesthetic score (and SD) by trial type on a 0 to 100 scale, with 0 being extremely aesthetically unpleasing, 100 being extremely aesthetically pleasing. Mean (and SD) ergonomic score by trial type

	Aesthetic score	Ergonomic score
IGAs Alone	80.4 (18.7)	20.85 (7.5)
IGA with GA as well	75.6 (20.7)	27.2 (4.26)

Principal Component Analysis did not yield usable results, and as such are not included here. The first principal component was heavily weighted by all variables except for Horizontal Button Spacing, and accounted for 33% of the overall variance. The second similarly did not show a clear variable contributing to the variance either, and did not (along with the other components), contribute to a negligible proportion of the overall variance.

Table 5.3. Mean (and SD) values of selected designs in the last generation by algorithm type.

	Button		Screen		
	Horizontal	Vertical	Vertical	Horizontal	Radius
IGAs Alone	6.62 (1.36)	3.32 (1.20)	35.91 (5.47)	43.40 (9.44)	7.38 (2.11)
IGA with GA as well	5.30 (1.10)	4.09 (0.74)	38.30 (3.05)	44.66 (5.33)	8.92 (1.23)

* Italics indicates pairs that are statistically significantly different ($\alpha = 0.0083$)

Discussion

When combining an IGA with a GA it would be easy to change the parameters in such a way that the combined algorithm does not still incorporate participant input, is deterministic, or does not converge. The results clearly show that these problems did not manifest themselves in this combined algorithm. The comparisons in Table 5.1 show that the algorithms were not deterministic.

When the mean value of a variable at the last generation is not the same as the mean of that variable's range, comparing the first and final generations' means serves as a good proxy for testing whether or not the algorithm is deterministic because the first generation is random, and therefore would be the same as the mean of the variable's range, and if the algorithm was deterministic the final generation would be the same as the first generation. By having final generation values that are not the mean of the variables', ranges we can be sure that the algorithm's starting state is not determining the final state because they are sufficiently different.

Along with testing if the algorithm is deterministic, it is important to test whether or not the algorithm converges. It was clear from subjective questionnaires that the algorithms converged (designs rated as "looking the same"). The data show that algorithm didn't just converge, but the algorithm converged to a solution that was highly desirable (one of the best designs they had seen). Additionally, the algorithm exceeded the user discriminability in the majority of the trials (all of the designs looked the same by the end). Overall, the combined IGA GA algorithm was worked extremely smoothly.

After verifying that the mechanics of the algorithm worked, it is important to see if the algorithm achieved the goals of the study; to make a difference in the final design when combined with an ergonomic rater. The fact that most of the independent variables were different when the GA teammate was present makes it clear that the GA worked in conjunction with the user to create more ergonomic designs.

The multi-objective equation used in the GA meant that some of the independent variables were "pushed" towards the maximum of their range by "larger-the-better" equations, and some were not. The two variables that were not significantly different were the two that were associated with the "middle-the-best" design goal of the GA, horizontal button spacing and screen width. Had all of the independent variables been significantly different between the algorithms, or had they all remained the same, it would be harder to tell if the combined algorithm was appropriately weighting the potentially competing values of its two inputs. Similarly, the lack of significantly lower aesthetics ratings when the GA was combined with the IGA also indicates that the GA did not overpower the user in sharing responsibility for the final design. Overall, the combined algorithm pushed designs to be taller, and have more rounded edges.

The physical ergonomic model used by the GA is a simplified multi-objective algorithm for the purpose of testing the concept of integrating aesthetic and usability factors rather than for its realism. Building a combined IGA and GA algorithm for ergonomics has not been done before. It is therefore important to test the concept in domains like mobile phones where ergonomics and aesthetics are well studied. It is also important to extend this work into new domains, especially those where user design preference is not as strong.

Chapter Five, Experiment Two

Experiment One tested aesthetics and physical ergonomics for mobile phones. To validate these findings and extend the work into the medical domain, the goal of Experiment Two is to replicate these findings in with blood glucose meters.

Method

Experiment Two extends Experiment One into the medical device domain. Experiment Two’s method was almost identical to Experiment One, except that in Experiment Two participants developed blood glucose meters.

Participants

All of the participants, except for one were engineering students at the University of Michigan. Of the 26 students, half were male, and half were female. The mean age was 20 years, with a standard deviation of 1.5 years. Participants used a wide variety of mobile phones; the most common was the Apple iPhone 4 (seven participants), followed by the Apple iPhone 3 (five participants). One of the participants had blood glucose meter experience, and used a blood glucose meter regularly on themselves, the rest had none. None of the participants had participated in prior IGA studies. The inclusion criteria were ownership or extensive use of a mobile phone, access to a computer with a high-speed internet connection and 19-inch to 21-inch monitor, and a lack of disabilities that would prevent them from safely using a computer for the duration of the study.

Variables

The variables in Experiment Two are the same as Experiment One. The IGA in Experiment Two used the same independent variables as the previous study, namely; horizontal screen size, vertical screen size, horizontal button spacing, vertical button spacing, and the radius of the device and the screen. The independent variables are shown in Figure 5.3. The blood glucose meter design used quadratic Bézier curves between the corners and points highlighted with stars in Figure 5.3 to draw the shell around the screen and buttons.

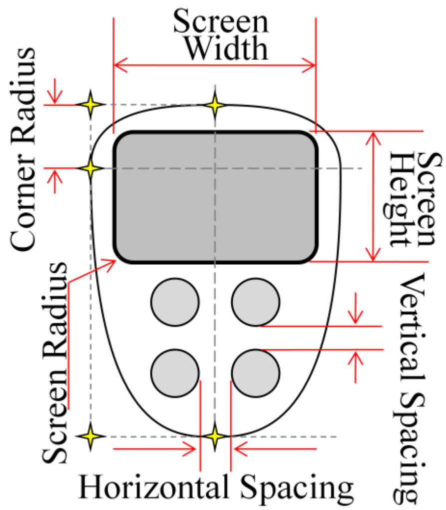


Figure 5.3. The blood glucose meter independent variables.

Algorithm Configuration

The algorithm used in this experiment was configured in the same manner as Experiment One in this chapter. The mutation rate, roulette wheel allocation, and genetic algorithm scoring method were kept the same.

Procedure

Participants were directed to the experiment website via a recruitment email to university department list-serves. Interested participants were qualified via a screening form on the experiment website. Once qualified, participants received information, filled out a consent form, received training, and completed the study remotely.

After consent, participants were given instructions about the tool, blood glucose meters, and how to complete the study. After training, participants completed a series of trials with a short questionnaire after each trial, and a longer debriefing survey after the last trial.

Participants completed ten trials; two practice, and eight test trials. Participants completed all of the trials alternating between trials with a GA input to the IGA, and IGA only trials. After all of the participants had completed the study, the data were analyzed in the same manner as in Experiment One.

Results

Participant trials were split by the presence of a Genetic Algorithm as well as an Interactive Genetic Algorithm; half of the trials had only the IGA (n=88), while the other half had an IGA and GA functioning at the same time (n=88). Of the trials rated (n=169), 87.0% said that their "final selection among the best presented in this trial." Additionally, participants rated 79.9% (n=135) of the trials "started looking the same to you." On the trials where it did start looking the same, the mean was trial 6.9 of the 10 trials.

Comparing the first and last generations through paired t-tests, in Table 5.1, shows that they are statistically significantly different. Ergonomic score was not tested in this comparison as it was not relevant. Similarly, Table 5.1 shows that the designs created using an IGA alone are statistically significantly different from designs with a Genetic Algorithm as well for most variables, including ergonomic score.

Table 5.4. Paired t-test *p*-values comparing the first and last generations, and the difference between IGAs alone and IGAs with GAs as well.

	Button		Screen		Radius	Ergonomic
	Horizontal	Vertical	Vertical	Horizontal		Score
First to last generation*	0.4709	1.954e-12	<2.2e-16	0.01468	<2.2e-16	
IGAs compared to IGAs with a GA+	0.09257	9.719e-5	1.181e-15	0.0003994	1.498e-06	3.934e-14

*$\alpha = 0.01$ after Bonferroni correction. $^{+}\alpha = 0.00833$ after Bonferroni correction.

The presence of a Genetic Algorithm did not statistically significantly alter the aesthetic score of the designs (*p*-value 0.676), as shown in Table 5.5. It did change the ergonomic score, *p*-value 3.934e-14, also shown in Table 5.5. The difference in mean values can be seen in Table 5.6.

Table 5.5. Mean aesthetic score (and SD) by trial type on a 0 to 100 scale, with 0 being extremely aesthetically unpleasing, 100 being extremely aesthetically pleasing. Mean (and SD) ergonomic score by trial type

	Aesthetic score	Ergonomic score
IGAs Alone	86.0 (15.1)	19.9 (7.4)
IGA with GA as well	85.1 (13.7)	26.7 (4.6)

Principal Component Analysis did not yield usable results, and as such are not included here. The first principal component was heavily weighted by all variables, and accounted for 35% of the overall variance. The second similarly did not show a clear variable contributing to the variance either, and did not (along with the other components), contribute to a negligible proportion of the overall variance.

Table 5.6. Mean (and SD) values of selected designs in the last generation by algorithm type. Pairs that were statistically significantly different ($\alpha = 0.0083$) are underlined.

	Button		Screen		
	Horizontal	Vertical	Vertical	Horizontal	Radius
IGAs Alone	6.18 (2.07)	<u>6.79</u> (2.14)	<u>29.07</u> (4.62)	<u>32.00</u> (7.49)	<u>9.35</u> (2.98)
IGA with GA as well	6.61 (1.78)	<u>7.95</u> (1.51)	<u>32.81</u> (2.66)	<u>34.56</u> (3.74)	<u>10.79</u> (1.55)

Discussion

The medical domain is a new area for a combined IGA GA algorithm testing. It also was used to test the combined algorithm in an area where users did not have experience using the device before. By using participants with little or no experience with blood glucose meters, Experiment Two reflects the experience that many newly diagnosed type two diabetics have when picking out their first blood glucose meter. They are often given the choice between several blood glucose meter designs, and then trained on its use.

The results of Experiment Two agree with Experiment One. The only notable difference is the difference in statistical significance between the algorithms for the screen width variable. Experiment One did not find the screen width to be significantly different between algorithm types, while Experiment Two did. This difference is likely an artifact of the way that the blood glucose meter calculates its shape, rather than a statement about user experience on design breadth.

General Discussion

In summary, these studies show that Interactive Genetic Algorithms combined with Genetic Algorithms are a valid method for user preference studies. Further, the studies demonstrate that IGAs combined with GAs can be used for product design for existing products and new product categories (blood glucose meters were new to the participants). Experiments One and Two used proven Interactive Genetic Algorithms to combine with a simplified, research-based, physical ergonomic teammate in the form of a Genetic Algorithm to generate designs that were aesthetically pleasing and safer to use. The primary difference between these studies and other researchers' work is the focus on aesthetics and ergonomics instead of less concrete terms such as "liking" in Brintrup, Ramsden, Takagi, & Tiwari (2008).

Combined, the experiments in this chapter show something important; users aesthetic preference is not changed as much with the presence of a GA teammate for mobile phones, devices they are familiar with, than blood glucose meters, devices they are not as familiar with. Comparing the shift in aesthetic score of designs with and without the GA in the two studies, aesthetic scores go down dramatically for mobile phones designed with the GA, but barely change at all for blood glucose meters designed with the GA. This may mean that IGA-GA combinations excel at incorporating design constraints for new products or with users who are unfamiliar with the product category.

Using IGA-GA combinations has the potential to have a large impact in the medical device industry. Combining IGAs with multi-objective GAs can make a design tool that can be used to help design medical devices so that they are safe and aesthetically pleasing to the patient. Design constraints, such as those outlined in AAMI/HE, Human Factors Engineering Committee (2010), or Kaye, North, & Peterson (2003), will need to be considered as part of the multi-objective design constraints enforced through the GA. Future work on the types of constraints that a GA can provide in conjunction with an IGA beyond the two types (middle-the-best, and larger-the-better) tried here will be important.

It would be interesting to see how a study similar to Seva, Duh, & Helander (2006), which tested emotional response to a number of mobile phones, would play out for blood glucose meters with novice users. Seva, Du & Helander (2006) found “intense emotional experience” from subjects, but this was may be connected to their familiarity with mobile phones, and may not exist for a new device they have not used before.

In the first study, screen width and horizontal button spacing were not significantly different with the presence of the Genetic Algorithm, but in the second study screen width was. The similarities and differences between devices has been tested (Chapter 4), and it was found that the preferred screen width and screen height:width ratio of phones and glucose meters were not significantly different. This, combined with the change in significance is important because it is a sign that users were allowing the GA to have a stronger effect on design preference between the devices.

The physical ergonomics GA used in both studies is an incredibly simplified model of physical ergonomics. The model is used here illustratively to test the ability of the combined algorithm to enhance the safety of the final design. Even though the algorithm is rudimentary, the resultant designs are likely more physically ergonomic.

Principal Component Analysis has been used in IGA studies to determine the variables where user opinion varies the most (Chapter 2). This can be useful to stratify potential designs, it is only useful when PCA finds a significant amount of the variance is derived by a small number of variables. PCA can be used in IGA, but it is not a robust solution that works in all cases. The reason that PCA may not work is because the variables are all important in creating variance. To combat this, the user group could be further segmented before PCA is done. With Apple iPhone participants for example, PCA may have found which independent variables contribute to the most variance. The large number of participants who used Apple iPhones in this study may have skewed the data. Based on previous research in Chapter 3, a more likely skew to the results would come from whether or not the phone that the user owned was a touchscreen phone (Chapter 3). This field of research would benefit greatly if a substitute for Principle Component Analysis (PCA) was be found to determine how to stratify multi-user feedback.

In the future, more work should be done to refine the multi-objective fitness function used by Genetic Algorithms. Additional physical ergonomics research is needed to enhance the realism of the designed algorithm. The multi-objective fitness function can also be used to include design constraints for the designs, such as required design envelope size.

References

AAMI/HE, Human Factors Engineering Committee. (2010). Human factors engineering - Design of medical devices. *ANSI/AAMI HE75:2009* . ANSI/AAMI.

Brintrup, A. M., Ramsden, J., Takagi, H., & Tiwari, A. (2008). Ergonomic Chair Design by Fusing Qualitative and Quantitative Criteria Using Interactive Genetic Algorithms. *Evolutionary Computation, IEEE Transactions on* , 12, 343 -354.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S. W., & Kim, J. (2004). Identifying Mobile Phone Design Features Critical to User Satisfaction. *Human Factors and Ergonomics in Manufacturing* , 14, 15-29.

Kaye, R. D., North, R. A., & Peterson, M. K. (2003). *Upcare: An Analysis, Description, And Educational Tool for Medical Device Use Problems*. Rockville, MD: Food and Drug Administration.

Nathan-Roberts, D., & Liu, Y. (2010). Investigation of Relative Mobile Phone Size Preference Using Interactive Genetic Algorithms. *54th Annual Meeting of the Human Factors and Ergonomics Society*. San Francisco, California: HFES.

Nathan-Roberts, D., Beeker, A., & Liu, Y. (2009). Modeling Two Key Physical Ergonomic Problems with Mobile Phones. *University of Michigan Engineering Graduate Symposium*. Ann Arbor.

Nathan-Roberts, D., Kelly, J. C., & Liu, Y. (2011). Determining the Effect of Users' Mobile Phone on Design Preference via Interactive Genetic Algorithms. *55th Annual Meeting of the Human Factors and Ergonomics Society*. Las Vegas, Nevada: HFES.

Nathan-Roberts, D., Liu, Y., Young, M., Hays, J. D., An So, C., Little, C., et al. (2009). Content Analysis of Literature on Mobile Phone Design Features. *17th World Congress on Ergonomics*. Beijing, China: IEA.

Seva, R. R., Duh, H., & Helander, M. G. (2006). Factors Affecting Emotional Responses in Mobile Phone Design Evaluation. *Human Factors and Ergonomics Society 50th Annual Meeting,* (pp. 2173-2177). San Francisco.

Chapter 6
Summary of Findings

The work reported in this dissertation help enrich our knowledge of aesthetic and ergonomic design in general and mobile device design in particular. It further demonstrate the value of IGA for aesthetic design and the use of IGA and GA together in holistic design integrating aesthetic and usability factors.

Summary of Aesthetics Findings

Within aesthetics, previous chapters outline how designers can modulate preference, shed light on cross-domain ties, and have specific takeaways for mobile device design.

How to Modulate Aesthetic Preference

The experiments in Chapter Three and Chapter Five show that preference can be modulated by changing the goals users have, the function of the device, or their experience.

Goals

The first study in Chapter Three shows that design preference can be changed by changing the user goal (specifically designing for one of three goals: aesthetically pleasing, functional, or both aesthetically pleasing and functional). It is therefore important for designers to consider the goals they give their participants carefully. Misrepresenting a product, or misunderstanding the goals of their customers could have severe negative consequences. User goals for a device often extend beyond the primary function of the device. For example, mobile phones are more than the purely utilitarian communication device that Martin Cooper, the inventor of the mobile phone, first used in 1973. Instead, mobile phones are multi-use devices that also have a significant aesthetic component (McMullan & Richardson, 2006). Modulating the goals of a user through suggestion or advertising, may also be a way to change a user's aesthetic preference by changing their goal. For example, if the calendar and meal recording features of a blood glucose meter are stressed in advertising, it may change the goals of the user, subsequently changing their aesthetic preference for that blood glucose meter.

Chapter Three, Study One goes on to show that users can combine multiple goals, each having an impact on design, or keep them separate. While users can hold multiple goals separate in their minds, they are not perfect at it. The prompt in Chapter Three, Study Two asked participants to focus only on designs only for dialing a phone number, but the distinct difference in outcomes makes it clear that it was difficult for users to ignore everything else they knew about mobile phones and how to use them. It is important for designers to keep in mind that participants will have a difficult time perfectly segmenting one task from another in aesthetics testing.

Function

In Chapter Three, Study Two, the function of a device (in this case a mobile phone that either has a touchscreen, or a built-in keyboard) has an effect on aesthetic preference. Even though the two devices looked identical, the way they functioned was different, and this changed participant aesthetic. This change in preference is also tied to users holding goals separate in their mind, as mentioned previously. It may have been hard for users to hold the other uses of mobile phones separate from using a phone only to dial numbers as they were prompted.

Ownership and Experience

Ownership (Chapter Three, Study Two), and experience (Chapter Five) play a role in aesthetic preference. Ownership was shown to modify aesthetic preference for mobile phones based on touchscreen or non-touchscreen phone ownership. Similar to ownership, experience played a role in Chapter Five. Experience may serve as a surrogate in some cases for ownership.

In Chapter Three, Study Two, touchscreen phone owners designed touchscreen and non-touchscreen phones with significantly taller screens than non-touchscreen phone owners. This may have been because of their experience of the benefits of a taller screen, something that touchscreen phones can easily provide because the keypads are minimized when not needed.

In Chapter Five, comparing Study One and Study Two, the degree to which aesthetic rating was modified by the presence of the Genetic Algorithm was very different. The presence of the Genetic Algorithm diluted the user's control over the final design. Participants rated the aesthetics of mobile phones almost 5 points lower on average when the GA was present, but only 0.9 points for blood glucose meters (p-values of 0.137 and 0.676 respectively). This is important because all of the mobile phone study participants had experience using mobile phones, but almost none (3 of 36) of the blood glucose meter study participants had experience using blood glucose meters.

Combined, these studies point to an effect of experience and ownership on aesthetic preference. It is important for designers to educate their users on how a device will be used when designing a new product, to mitigate the effect of a lack of experience.

<u>Having Control Over the Design</u>

In Chapter Five, Study One, participants created mobile phones with or without an ergonomic Genetic Algorithm as a teammate. The designs with a teammate were significantly different for three of the five independent variables in the mobile phone design, but the rating of aesthetics was not significantly different. In Chapter Five, Study Two, participant designed blood glucose meters were significantly different when the ergonomic teammate was present for four of the five independent variables in the design, but the aesthetic ratings varied even less between designs with an ergonomic genetic algorithm and without. One of the most important takeaways of this research on how to modulate aesthetic preference is a clear example of cognitive dissonance; that designing a device yourself will increase your preference for it.

Cross-Domain Ties of Aesthetic Preference Findings

Some aesthetic preference similarities exist across the domains of mobile phones and handheld blood glucose meters. Chapter Four and Chapter Five test the aesthetic ergonomic connections between these devices. Mobile phones and blood glucose meters are similar in that they are both handheld digital input/output devices, which makes comparing them easier than more dissimilar objects. Chapter Four shows that the fidelity of the designs, realism, and their similarity effect the cross-domain connections. A summary of the cross-domain findings in Chapter Four is presented in Table 6.1. Chapter Five re-affirms these connections by finding similar independent variables changed by the Genetic Algorithm for both devices. The most important things for designers to consider when looking across domains for design cues are that the exact values are not as informative as the ratios of design variables to each other, and that the fidelity, realism, and similarity of designs are important to consider when doing direct cross-domain comparisons.

Table 6.1. Chapter Four outcomes highlighting similarities and differences across domains

Variable	**Study 1** *Phone to meter*	**Study 2** *Previous Meter experience*	**Study 2** *Phone to Meter A*	**Study 2** *Phone to Meter B*	**Study 2** *Phone to Meter C*
Button W spacing	Different	Different for phones*	Different	Different	Different
Button V spacing	Different	Similar	Different	Different	Different
Screen V	Different	Meters A & B different*	Different	Different	Different
Screen W	Similar	Meter A & C different*	Different	Different	Different
Radius	Similar	Different for phones*	Different	Similar	Similar
Screen area to key area ratio	Similar	Meter A & C different*	Different	Different	Different
Screen V to screen W ratio	Similar	Meter C different*	Similar	Similar	Similar
Screen W to Key W ratio	Different	Meter C different*	Similar	Similar	Similar
Screen V to KeyV ratio	Different	Meter C different*	Similar	Similar	Similar
Keypad V to keypad W ratio	Different	Different	Different	Similar	Different
Overall height to width ratio	Different	Different	Different	Different	Similar

* Denotes statistically significantly different before correction for multiple tests

Summary of Mobile Device Specific Aesthetic Findings

The two mobile device specific aesthetics takeaways are that individual components have a large impact on aesthetic preference, and that the classic definitions of beauty serve as a rough guide for design, but do not apply perfectly. Content analysis in Chapter Two found that each component's size and their relative sizes played a strong part in the aesthetic of mobile phones, more so than just the overall shape. Testing of common aesthetics ratios for mobile phones and blood glucose meters was performed in Chapter Four. These tests showed that the aesthetics ratios were similar to, but not the same as established values (Marcus, 1992). A summary of all of the aesthetics findings is shown in Table 6.2.

Table 6.2. Research outcomes relating to aesthetics.

Modulating Preference	***Study***
People can hold goals separately and combine them	*Chapter 3, Study 1*
Goal affects preference (aesthetics, function, both)	*Chapter 3, Study 1*
Users can not perfectly separate or ignore goals, when instructed	*Chapter 3, Study 2*
Function changes aesthetic preference (touchscreen vs. non-touch)	*Chapter 3, Study 2*
Ownership has an effect on preference (touchscreen phones)	*Chapter 3, Study 2*
Lack of experience decreases the strength our aesthetic preference	*Chapter 5, Study 2*
Cross-domain ties	
Some similarities exist across domains	*Chapter 4, Study 1*
Some, but not all aesthetic ratios are similar across designs	*Chapter 4, Study 2*
Fidelity, realism, similarity effect cross-domain connection	*Chapter 4, Study 2*
Aesthetics not affected by GA	*Chapter 5, Study 1*
Mobile Devices	
Component size has a large impact on preference,	*Content Analysis*
Compared to classic types of beauty: Close but not good fit	*Chapter 1, Study 1*

Summary of Physical Ergonomics Findings

Genetic Algorithms

Chapter Five showed that a Genetic Algorithm can enhance the physical ergonomics of a mobile device being designed with an Interactive Genetic Algorithm. The Genetic Algorithm was able to use a multi-objective physical ergonomics model as the fitness function it used to rate the devices. The algorithm made both mobile phone and blood glucose meters significantly more ergonomic, but did not significantly decrease the user aesthetic rating. The implication of these findings is that designers can use Genetic Algorithms in conjunction with user-driven Interactive Genetic Algorithms to optimize a number of other aspects of a design, such as physical ergonomics and engineering design constraints simultaneously. A summary of the physical ergonomics findings of this work are shown in Table 6.3.

Summary of Mobile Device Specific Physical Ergonomics Findings

Little device specific research exists on the physical ergonomics of mobile phones, and even less on blood glucose meters (Croasmun, 2004). Chapter Two outlines several ergonomic concerns specific to mobile devices. Specifically, screen size is important to legibility, and increasing phone length can potentially reduce the likelihood or effects of Cubital Tunnel Syndrome by reducing elbow flexion. Table 6.3 outlines the research findings for physical ergonomics.

Table 6.3. Research outcomes relating to physical ergonomics.

Genetic Algorithms (Chapter Five)
Genetic Algorithms can increase the physical ergonomics
Aesthetic rating by the user is not necessarily effected by the GA
Lack of user experience with a device decreases the strength of user aesthetic preference
Mobile device specific (Ergonomics models, Chapter Two)
Screen size is important for legibility, and is often too small in mobile devices
Overall length matters for mobile phones

Summary of Findings Related to Aesthetic Ergonomics Research Methods

Several concepts that are relevant to aesthetic ergonomic research methods have already been covered in this chapter. Specifically, that the Genetic Algorithm can positively impact the ergonomics of the device design without significantly negatively impacting the aesthetics, and that device experience can play a role in participant designs. It is important for designers to be aware that experience with the IGA interface can be enhanced by providing a suitable number of practice trials before test trials.

Summary of Interactive Genetic Algorithm Specific Findings

Chapters Two and Four show that Interactive Genetic Algorithms work in these two domains to allow users to design devices they find aesthetically pleasing. Chapter Five found that the Interactive Genetic Algorithm can be effectively combined with a multi-objective Genetic Algorithm to provide an ergonomic teammate to users designing for aesthetics. The algorithm is setup in such a way that it converges to a solution, is non-deterministic, and exceeds participant’s threshold of perception of the differences between the designs. Chapter Two found that Principal Component Analysis can be used to segment users. While this was not found to work in later studies, it may be possible to segment users into smaller categories (such as touchscreen smartphone users) that would allow the PCA to provide more feedback. Chapter Three, Study One also found that participant location did not affect the results of these studies allowing the remaining studies to be completed remotely. The results mentioned above as well as other relevant findings mentioned previously in this chapter are shown in Table 6.4.

Table 6.4. Research outcomes relating to Aesthetic Ergonomic Research Methods

Finding	***Study***
IGAs work in these domains	*Chapters 2 & 4*
Fidelity, realism, similarity affect cross-domain connection	*Chapter 4, Study 2*
Principal Component Analysis can sometimes segment users	*Chapter 2, IGA*
Genetic Algorithms can be used as effective ergonomic teammates	*Chapter 5*
Genetic Algorithms do not necessarily affect Aesthetic ratings	*Chapter 5*
IGA studies do not need to be performed locally	*Chapter 3, Study 1*
Practice trials are important	*Chapters 4 and 5*

Future Directions

Future Aesthetics Research Directions

Several areas of aesthetics touched upon in this dissertation merit further study. Specifically the strength of our aesthetic preference for devices we design, the effect of experience, and user ability to hold multiple goals separate warrant further study. Users showed a strong preference for the devices that they designed. Using methods like the Technology Acceptance Model (Szajna, 1996), it would be very helpful to test the degree to which self-designed medical devices can reduce device abandonment. Device abandonment is a serious issue in the rapidly growing home healthcare field. Self-designed medical devices could potentially have a huge impact on the medical outcomes.

Another interesting test of the strength of our preference for our own designs would be to test that strength in the presence of peer designed devices. This could include having participants design a device, and then being shown their design among other designs, and asking for their favorite. The experimenter could change the aesthetic nature of the cohort designs, and the background that they provide about those devices. The user device could be compared to a set of devices previously rated as not appealing by other users, or as appealing by other users. The experimenter could inform the participant that the other designs were designed by famous designers, their peers, or a less desirable peer group such as school children.

The effect of user experience was tested in Chapters Four and Five. Research found that participants designing a product they had little or no experience with had a more moldable sense of aesthetics (the Genetic Algorithm affected their final designs more). Research to achieve better understanding of how best to teach someone about a new domain so that they can have a strong sense of aesthetic preference within that domain would be very useful for product designers. By testing various training methods, researchers could make aesthetic ergonomics more applicable to use in new products and other domains where users have little experience.

The first study in Chapter Three tested our ability to separate multiple goals. Further research testing how many goals users can distinctly hold in mind would aid designers who have complex design problems, such as vehicle design, where IGAs have been used to design aerodynamic vehicles (Kelly, Maheut, Petiot, & Papalambros, 2011).

Study Two, Chapter Three suggested that users may have a difficulty separating their other knowledge about a device domain in general from the specific capabilities and uses of a particular device. This can cause problems for designers designing new products, or designing products which have a specific, niche use. Testing how prompt and other coaching can be used to better encourage participants to separate their knowledge about other devices from their current device would be very useful to designers.

Future Physical Ergonomics Research

Future directions for the physical ergonomics aspects of this research include building a stronger, biomechanics based ergonomics rating tool. Such a tool could use actual-size devices and anthropometric measurements of participants.

Future Tools and Methods

This dissertation was not a dissertation about genetic algorithm optimization, and as such did not explore a number of the interesting questions and methods uncovered in this area. Use of new technology, like HTML5, as well as new computational methods should be explored to further optimize the genetic algorithms used within this dissertation.

Several exciting research areas exist in building tools and methods for use in aesthetic ergonomics research. It would be useful to increase the fidelity and range of options available for product design with Interactive Genetic Algorithms. Such improvements could include the implementation of non-continuous variables, such as a mobile phone which either has a full keyboard or a small number pad, or has a feature, such as a camera, which is either available or not depending on previous selections. These improvements could improve how interesting participants found the experiments which may help IGA adoption for end-user-designed product development.

This dissertation was also not a treatise in statistical methods. It is probable that better statistical methodologies exist, and may even be used for IGA analysis. The methods used here were sufficient to these problems, and likely did not alter the findings, but further statistical method development may be warranted. To further improve the IGA, a suitable replacement to Principal Component Analysis would allow for effective segmentation of the user population.

To enhance the effectiveness of the Genetic Algorithm teammate used in Chapter Five, it would be useful to develop more complex multi-objective fitness. A more complex fitness function could include hard-limits, like minimum envelope size, or dynamically change the amount of control it has over future generations as it approaches solutions that are above a specific threshold fitness score.

To test device abandonment, as suggested above, it would be important to be able to build physical models of the devices designed by participants. Combining an Interactive Genetic Algorithm with new, low cost rapid prototyping technology would open up endless possibilities for designers.

Using these physical prototypes based on rapid prototyping technology, or building computational models of user task performance to determine how fast a task can be completed with a design is, or its error rate would help designers immensely.

Conclusion

This dissertation work helps us to better understand the workings of user aesthetic preference, and the use of genetic algorithms in the nascent field of aesthetic ergonomic design. There are a number of very exciting areas of future research which include using rapid prototyping and Genetic Algorithms to test user-designed medical device acceptance.

References

Croasmun, J. (2004, July 16). Are Ergonomists Really Consulted in Mobile Phone Design? *Ergonomics Today (Quick news--Open Access)* . ErgoWeb.

Kelly, J. C., Maheut, P., Petiot, J.-F., & Papalambros, P. Y. (2011). Incorporating user shape preference in engineering design optimization. *Journal of Engineering Design* , 22, 627--650.

Marcus, A. (1992). *Graphic Design for Electronic Documents and User Interfaces*. New York: ACM Press.

McMullan, J., & Richardson, I. (2006). The Mobile Phone: a hybrid multi-platform medium. *Proceedings of the 3rd Australasian conference on Interactive entertainment*. Australia: Murdoch University.

Szajna, B. (1996). Empirical Evaluation of the Revised Technology Acceptance Model. *Management Science* , 42, pp. 85--92.