May 2020

Using a Bayesian Framework to Develop 3D Gestural Input Systems Based on Expertise and Exposure in Anesthesia

Katherina Ann Jurewicz
Clemson University, kjurewi@g.clemson.edu

USING A BAYESIAN FRAMEWORK TO DEVELOP 3D GESTURAL INPUT SYSTEMS BASED ON EXPERTISE AND EXPOSURE IN ANESTHESIA

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Engineering

---

by
Katherina Jurewicz
May 2020

Accepted by:
Dr. David M. Neyens, Committee Chair
Dr. Ken Catchpole
Dr. Mary Beth Kurz
Dr. Jamiahus Walton

ABSTRACT

Interactions with a keyboard and mouse fall short of human capabilities and what is lacking in the technological revolution is a surge of new and natural ways of interacting with computers. In-air gestures are a promising input modality as they are expressive, easy to use, quick to use, and natural for users. It is known that gestural systems should be developed within a particular context as gesture choice is dependent on the context; however, there is little research investigating other individual factors which may influence gesture choice such as expertise and exposure. Anesthesia providers' hands have been linked to bacterial transmission; therefore, this research investigates the context of gestural technology for anesthetic task. The objective of this research is to understand how expertise and exposure influence gestural behavior and to develop Bayesian statistical models that can accurately predict how users would choose intuitive gestures in anesthesia based on expertise and exposure.

Expertise and exposure may influence gesture responses for individuals; however, there is limited to no work investigating how these factors influence intuitive gesture choice and how to use this information to predict intuitive gestures to be used in system design. If researchers can capture users' gesture variability within a particular context based on expertise and exposure, then statistical models can be developed to predict how users may gesturally respond to a computer system and use those predictions to design a gestural system which anticipates a user's response and thus affords intuitiveness to multiple user groups. This allows designers to more completely understand the end user and implement intuitive gesture systems that are based on expected natural responses.

Ultimately, this dissertation seeks to investigate the human factors challenges associated with gestural system development within a specific context and to offer statistical approaches to understanding and predicting human behavior in a gestural system.

Two experimental studies and two Bayesian analyses were completed in this dissertation. The first experimental study investigated the effect of expertise within the context of anesthesiology. The main finding of this study was that domain expertise is influential when developing 3D gestural systems as novices and experts differ in terms of intuitive gesture-function mappings as well as reaction times to generate an intuitive mapping. The second study investigated the effect of exposure for controlling a computer-based presentation and found that there is a learning effect of gestural control in that participants were significantly faster at generating intuitive mappings as they gained exposure with the system. The two Bayesian analyses were in the form of Bayesian multinomial logistic regression models where intuitive gesture choice was predicted based on the contextual task and either expertise or exposure. The Bayesian analyses generated posterior predictive probabilities for all combinations of task, expertise level, and exposure level and showed that gesture choice can be predicted to some degree. This work provides further insights into how 3D gestural input systems should be designed and how Bayesian statistics can be used to model human behavior.

ACKNOWLEDGMENTS

There are several people who have helped and supported me throughout my dissertation studies, and I am forever grateful for the support I received. My advisor, Dr. David Neyens, was important to the completion of my dissertation studies. He is the best advisor (not an exaggeration), and I'm thankful for him pushing me to be a better researcher, teacher, and colleague. He has inspired me to pursue a career in academia, and I hope to be as successful and influential as him some day. I would also like to thank my other committee members: Dr. Ken Catchpole, Dr. Mary Beth Kurz, and Dr. Jamiahus Walton. I am grateful for their time, effort, and advice they provided to me throughout my studies.

I would also like to recognize my labmates, current and former, in the Ergonomics and Applied Statistics Lab. The EASt lab is a special place where no idea is bad, no person is judged, and we all help each other be better researchers, teachers, and people. I'd like to thank all of my current and former labmates for all of the help and encouragement during our time working together. I would also like to thank all of the faculty and staff in the Department of Industrial Engineering at Clemson University for your support through my years at Clemson.

I would also like to thank my family and friends both in and outside of Clemson - I'm especially thankful for my Clemson family and my graduate school family for making my graduate school experience one I will always remember.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: OVERVIEW

Computers have become closely integrated in our homes, cars, and work environments and cannot be easily separated from our daily routines. Although the keyboard and mouse have dominated the human-computer interaction (HCI) market (Hinckley, Jacob, Ware, Wobbrock, & Wigdor, 2014), the expanse of possible interactions with a keyboard and mouse falls short of human capabilities. What is lacking in the technological revolution is a surge of new and natural ways of interacting with computers (Hinckley et al., 2014). In-air gestures are a promising input modality as they are highly expressive, easy to use, quick to use, and natural for users. Gestural interaction is unique because the human is the input device, simplifying the transfer and information between human and computer (Hinckley et al., 2014).

Despite the naturalness and ease of use of gestures, 3D gestural technology is not as prevalent in HCI as other input modalities. There are a variety of factors hindering the advancement of gestural control in HCI from both a technical perspective of developing robust and reliable technologies as well as from a human factors perspective of creating a natural and intuitive user experience. Gestures as an input modality is a unique design problem because it is beneficial for the user to be able to use natural hand gestures, but it is difficult to design intuitive gestural systems across all users because gestures are typically highly individualized (Stern, Wachs, & Edan, 2008). It is known that gestures should be developed within a particular context (Ardito, Costabile, & Jetter, 2014; Jacob & Wachs, 2014a; Jacob, Wachs, & Packer, 2013; Nielsen, Störring, Moeslund, &

Granum, 2004; Wigdor & Wixon, 2011), and all types of user groups, from novice to expert, should be able to interact with gestural systems naturally (Wigdor & Wixon, 2011). However, there is limited research investigating the effect of expertise on gesture choice (Jurewicz, Neyens, Catchpole, & Reeves, 2018). Furthermore, gestural systems are intended to be integrated into work environments for extended periods of time, but there are no studies which investigate the gesture choices of users over time, despite the skills-, rules-, and knowledge-based taxonomy demonstrating how human behavior may change depending on task demands and the level of cognitive control (Rasmussen, 1983). Therefore, this research seeks to investigate how expertise and exposure influence gesture choice under a specific context.

This dissertation additionally seeks to predict how users will respond gesturally in a 3D, vision-based gestural input system. If gesture choice can be predicted, then researchers may be able to design intuitive gesture systems that are based on expected natural responses. Gestures are difficult to learn (Hinckley et al., 2014), and there is little intuitive gestural agreement among users (Wobbrock, Morris, & Wilson, 2009). In other words, there is much uncertainty when it comes to gestural interaction from an observer's standpoint, thus it may be advantageous to quantify the uncertainty behind gestural HCI. A Bayesian framework may be appropriate for understanding novel input systems, such as integrating 3D gestural input into human-computer interaction, because Bayesian statistics provides a quantitative framework for modeling uncertainty. Bayesian frameworks are able to represent knowledge of the current state of the system and are able to integrate new information for understanding a new state of a system (Cowles, 2013). This dissertation integrates Bayesian statistics into the data analysis to predict gesture choice for HCI in the context of anesthesiology.

Research Objective and Questions

The objective of this research is to understand how expertise and exposure influence gestural behavior and to develop models that can accurately predict how users would choose intuitive gestures based on expertise and exposure. Expertise and exposure may influence gesture responses for individuals; however, there is limited to no work investigating how these factors influence intuitive gesture choice and how to use this information to predict intuitive gestures to be used in system design. Therefore, my dissertation research questions are:

1. How does expertise influence how users gesturally respond to a computer system?
2. How does exposure to the gestural system influence how users gesturally respond to a computer system?
3. How accurately can natural gestures be statistically predicted?

Answering these research questions will provide a more complete understanding of the end users to aid in developing intuitive gesture systems that are based on expected user responses.

CHAPTER TWO: LITERATURE REVIEW AND RESEARCH FRAMEWORK

Gestural Technology

Gestures are a readily available means of communication and are integral in communicating with other humans (Efron, 1941; Freedman, 1972; Kendon, 1988; McNeill, 1992). Gestures can support and even replace verbal communication when speech is hindered or impossible (McNeill, 1992). Gestures come so easily and natural to people that infants may use gestures to communicate before they learn to talk; gestures are used when travelling to foreign countries and communicating with someone who speaks a different language; and generally, people use gestures in any verbal conversation to support key ideas. Gestures can also be used when someone is too far away or a barrier occludes verbal communication, such as an air marshal who signals with their hands to help pilots navigate an airplane on a tarmac. Since gestures are a natural part of human-human communication, they can also serve as a natural way to communicate with computers and other devices (Karam & Schraefel, 2005), and with recent advancements in motion sensing technology, we introduce new avenues of human computer interaction and natural ways to interact with computers via gestures.

Human-computer gestures can be 2D or 3D, depending on the type of technology used for the gestural interaction. 2D gestures are primarily contact-based and performed on a touch-sensitive surface (e.g., touchscreen on a smartphone). 3D gestures are primarily vision-based where the gesture is captured by a camera system (e.g., Microsoft Kinect). 3D, vision-based gesture systems could additionally be based on either hand movement or body movement. For example, Microsoft Kinect is a 3D, vision-based gestural technology and is able to capture full body gestures whereas the Intel RealSense Camera is programmed to capture fine hand and finger movements. 3D, vision-based systems allow users to gesture in any direction, whereas 2D, contact-based systems, such as touchscreens, only allow users to perform static or dynamic gestures in either one or two directions. The use of a camera eliminates the need for additional sensors or other equipment (e.g., data glove) eliminating an unnatural and intrusive user experience (Baudel & Beaudouin-Lafon, 1993). Due to the nonintrusive component and the degree of flexibility, 3D, vision-based gestural systems have the potential to fill the need of creating natural and nonintrusive interaction experiences in HCI.

For HCI in general, there are traditionally two approaches to developing new computing systems: centering around the technology or centering around the abilities of the human (Hinckley et al., 2014). Gestural technology follows these same traditional forms of HCI development as researchers may either take a technology-based approach, focusing on maximum recognition accuracy of gestures, or take a human-based approach, focusing on maximizing usability of the system as shown in Figure 1 (Nielsen et al., 2004).

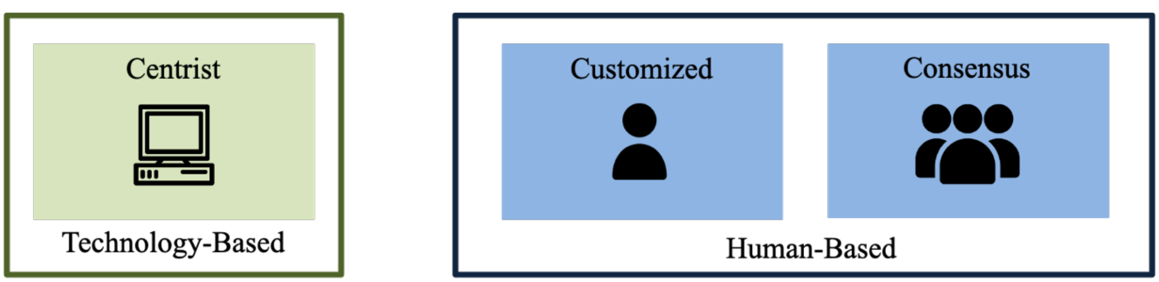


Figure 1. Depiction of human-based and technology-based gesture development

The technology-based approach is more recently known as the centrist approach and is when a research group or designers choose the gesture vocabulary of the system and teach the gesture commands to end users (Stern et al., 2008). The gesture vocabularies under a centrist approach are typically distinguishable gestures that can be recognized easily by the software (Nielsen et al., 2004; Stern et al., 2008); however, since the gestures are derived based on technology capabilities, these systems may be implemented at the expense of usability and user intuition. On the other hand, human-based approaches are centered around the user (Nielsen et al., 2004). Human-based approaches are broken down into a consensus approach or a customized approach (Stern et al., 2008). In both consensus and customized approaches, gestures are elicited directly from end users. In the consensus approach, users in an experimental group independently reach a "consensus" on which gestures intuitively map to the set of functions (i.e., the most frequently performed gesture for a function across a group is the most intuitive mapping for the function; Stern et al., 2008). The customized approach takes a more in-depth user elicitation approach where each user of the system defines their own gesture vocabulary (Stern et al., 2008). The customized approach is entirely user-centered as each user defines their own gestures, and the gestures derived are completely natural and intuitive to the specific user. However, the computational effort of a system based on the customized approach is very high, and as the user group population grows in size, the development of the system may not scale well. In both the consensus and customized approaches, users often choose gestures that are intuitive, natural, and comfortable, but the users may fail to consider technological limitations (Stern et al., 2008). Therefore, usability is maintained but possibly at the expense of recognition accuracy and subsequently reliability of the system.

The tradeoff between usability and accuracy is one of the primary roadblocks to gestural system success, and it is not explicitly clear which approach is more advantageous. Despite both technology-based and human-based approaches being used in practice, Morris, Wobbrock, and Wilson (2010) compared a gesture set elicited from users and a gesture set developed by HCI researchers and demonstrated that participants preferred user-defined gestures over researcher-defined gestures. This finding suggests that participatory design methodologies are critical when developing a gesture vocabulary (Morris et al., 2010) and advocates for the user-centered approach that human-based methods offer.

Applications of Gestural Technology

Several researchers have investigated ways to implement gestural technology in a range of applications. Before vision-based, 3D gestural input technology, 3D gestures were often captured via a glove worn on the hand. Charade is an example of this type of technology in which a data glove is worn and gestures are used to control a computer-based presentation (Baudel & Beaudouin-Lafon, 1993). However, glove-based devices proved to be intrusive since the user has to wear a glove, taking away from a natural communication experience (Sturman & Zeltzer, 1994). Aside from glove-based devices, television control by the means of a 3D, vision-based system was demonstrated to be possible, but the system only utilized the open-palm gesture (W. T. Freeman & Weissman, 1995). With the refinement of vision-based, 3D gestural input devices, there is the potential to expand gesture sets beyond one gesture for use in HCI. However, it is known that a gesture in one context may have a completely different meaning in a different context, thus gestural input systems need to be developed for a specific context and application (Nielsen et al., 2004). The system is not expected to succeed when there is not a specific application and general gesture vocabulary sets are used for HCI across multiple applications (Ardito et al., 2014).

Transportation is one domain which has seen considerable interest in developing gestural control in the vehicle. With increasing internet capabilities and cloud-based computing in vehicles, drivers' attention is demanded more than normal because of the potential to become engaged with secondary tasks (e.g., navigation, radio, mobile phones; Jæger, Skov, & Thomassen, 2008). Additionally, the recent push for semi-autonomous vehicles further provides the driver with possibilities for engagement in secondary tasks, thus leading to distracted driving and increased risk of crashes (Klauer, Guo, Sudweeks, & Dingus, 2010). With increases in the number of possible secondary tasks drivers can engage in, this comes with increased buttons, switches, and indicators surrounding the dashboard, thus increasing the visual workload of the driver (Riener et al., 2013). The controls can be visually overwhelming and the information that they represent must be processed concurrently with the primary driving task, despite the controls being designed to support the driver. Performing the secondary tasks with the controls while driving requires a high amount of perceptual, cognitive, and physical skill (Yale, Hansotia, Knapp, & Ehrfurth, 2003), and the secondary control tasks compete for the same perceptual and cognitive resources as the primary driving task (Normark, Tretten, &

Gärling, 2009). Driver distraction is already a leading cause for major crashes as drivers are frequently moving their attention away from the primary task of driving to secondary, non-driving activities (Klauer et al., 2010). Car manufacturers are continuously competing with the need to improve driver experience and the need to maintain a high level of safety. Since driving is primarily a visual task, it is important to study drivers' visual workload (Riener et al., 2013) and seek ways to keep drivers' eyes on the road in order to maintain or improve safety (Pickering, 2005).

3D, vision-based gestural technology is suitable for the transportation domain as it does not require visual attention which allows the driver to maintain their eyes on the road simultaneously manipulating controls in the vehicle. There has been a push to develop novel user interfaces, such as gestural control, that make it easy to operate in-vehicle systems while driving (Ashley, 2014). As of 2014, Audi, BMW, Cadillac, Ford, General Motors, Hyundai, Kia, Lexus, Mercedes-Benz, Nissan, Toyota, and Volkswagen are all working on incorporating gestures as an input mode for secondary tasks (Ashley, 2014). Gestural interfaces have been shown to reduce driver distraction (Ohn-Bar & Trivedi, 2014) and reduce visual and cognitive workload of the driver (Jæger et al., 2008; Riener, 2012) which ultimately may increase safety. There has been extensive research investigating the use of 3D, in-air gestures; however, a majority of these studies are technology-based or centrist-developed systems (Akyol, Canzler, Bengler, & Hahn, 2000; Cairnie, Ricketts, McKenna, & McAllister, 2000; Ohn-Bar & Trivedi, 2014; Parada-Loira, González-Agulla, & Alba-Castro, 2014; Rahman, Saboune, & El Saddik, 2011). Very few studies found in the literature review incorporated human-based methods for 3D gestural interfaces (Fariman, Alyamani, Kavakli, & Hamey, 2016; Riener & Rossbory, 2011) despite human-based methods being preferred over technology-based methods (Morris et al., 2010). Of the human-based systems, only a few climate control and some infotainment functions were investigated with limited results (Fariman et al., 2016; Riener & Rossbory, 2011).

As shown in this review of the literature for transportation gestural HCI, technology-based approaches are considerably more prevalent than human-based methods. The technology-based systems in transportation are potentially highly accurate at the expense of usability and intuition to the end user. (Nielsen et al., 2004; Stern et al., 2008). If this is the case and gestures are not intuitive and natural to the user, then the user must use cognitive resources to explicitly remember the gesture-function mapping thus potentially increasing mental workload, decreasing performance, and increasing driver distraction. Furthermore, if the gestural system is not intuitive and usable then incorrect gestures may be performed resulting in system errors, so regardless of the technology-based system being highly accurate, there might still be errors due to a lack of usability. System errors lead to user frustration, less trust in the system, and eventually disuse which leads to overall low adoption of the technology. We are seeing these effects in real time within the transportation domain and the manufacturers who are developing gestural systems quickly via a technology-based approach, and we can learn from transportation how to move forward with future development of gestural systems (Ashley, 2014).

Operating Room

A second domain that can benefit from 3D, vision-based gesture systems is healthcare, especially in areas which must stay clean like the operating room (OR). Healthcare has been on the forefront of investigating gestural systems because the touchless interactions are useful in preventing the spread of pathogens and preventing patients from contracting healthcare-associated infections (HAIs) subsequently preserving sterility in ORs and other clean environments (Wachs et al., 2008). Gestural control could be advantageous to several healthcare applications and environments as the U.S. Department of Health and Human services (2013) states that HAIs can be contracted anywhere across the continuum of care for a patient. In 2002, there were approximately 1.7 million cases of HAIs among U.S. Hospitals with 99,000 associated deaths (Klevens et al., 2007). It is also estimated that hospital-contracted HAIs account for $28 billion to $33 billion in healthcare costs every year (U.S. Department of Health and Human Services, 2013). As modern healthcare continues to increase in complexity, it is important to develop innovative ways to combat bacterial infection such as through technological interventions.

In 2010, The Society of Healthcare Epidemiology of America (SHEA) offered a national approach to HAIs and minimizing bacterial transmission (The Society of Healthcare Epidemiology of America, 2010). Since its release, numerous studies have strengthened the understanding of HAIs and developed prevention techniques to be implemented hospital-wide, including but not limited to further training (Barsuk, Cohen, Feinglass, McGaghie, & Wayne, 2009; Comer et al., 2011), improvement in hand hygiene (Pittet et al., 2000; Sax et al., 2007), and best practices guidelines for healthcare providers (Marschall et al., 2014). However, due to the nature of work in the OR (i.e., interaction with one patient over a long period of time), these measures may not be adequate to eliminate contamination (Stackhouse et al., 2011). Additionally, the surface environment has been extensively connected to HAIs (Weber, Anderson, & Rutala, 2013); pathogens can survive on hospital room surfaces and medical equipment for hours, days, and even up to months (Weber et al., 2013). As healthcare providers in the OR care for multiple patients while touching multiple surfaces and equipment, they are potentially facilitating the transfer of bacteria from one patient to another.

3D, vision-based gestural technology has been introduced to the healthcare domain focusing on minimizing bacterial spread in the OR. A majority of the healthcare literature focuses on surgeons interactions in the OR and navigating radiological images during a case (Bizzotto et al., 2014; Jacob & Wachs, 2014b; Jacob et al., 2013; Mewes, Saalfeld, Riabikin, Skalej, & Hansen, 2016; Schroder, Loftfield, Langmann, Frank, & Reithmeier, 2014; J. Wachs et al., 2006). Surgeons scrub into a surgical case and must stay sterile while working at the incision site, and if the surgeon wants to review radiological images, they must remove themselves from the sterile field to interact with non-sterile technology. This is a timely process, so it is beneficial for surgeons to be able to interact with medical imaging in a sterile manner. The studies investigating surgeon's interactions in the OR have shown positive results of the technology and positive feedback from the users. However, there are other providers in the OR who do not scrub into a surgical case, such as anesthesia providers, who may benefit from touchless HCI.

Anesthesia providers interact greatly with the patient before, during, and after a surgical case, and it has been shown that anesthesia providers switch tasks about every six seconds (Jurewicz et al., Under Review). If new hand-hygiene steps are introduced into the anesthetic task flow, this could potentially impact overall workflow and subsequently introduce new and unanticipated patient safety events. Thus, it is important to consider how bacterial contamination can be mitigated in a way that is cohesive to current anesthetic work. One way is integrating new technologies that support infection control and workflow such as touchless interactions via gestural control.

Anesthesia Workstation

There is plentiful evidence in the literature that shows that the anesthesia workstation is often contaminated and anesthesia providers are linked to bacterial transmission in patients. A novel study sought to understanding the dynamics of bacterial spread in the anesthesia workstation. They simulated the bacterial contamination in the anesthesia workstation by using fluorescent marker and having anesthesia providers perform the intubation process (the anesthetic step that occurs before the operation begins) as they normally would (Birnbach, Rosen, Fitzpatrick, Carling, & Munoz-Price, 2015). Although the fluorescent marker was initially present only inside the mouth and on the lips of the patient simulator, the fluorescent marker spread throughout the anesthesia environment during the intubation process (Birnbach et al., 2015). Thirteen areas within the anesthesia environment (including the IV hub, anesthesia machine surface, anesthesia circuit, oxygen valve, and anesthesia cart) were contaminated in 100%

of the observations, and the computer keyboard was contaminated in 80% of the observations (Birnbach et al., 2015). This study demonstrates that there is potential for widespread bacteria contamination before the operation even begins. A separate observational study showed that the anesthesia environment has bacterial transmission in 89% of the observed surgical cases (Loftus et al., 2011). These findings support the notion that there is a cyclical pattern of bacterial transmission from the patient to the anesthesia environment back to the patient and there is widespread bacterial contamination in the anesthesia workstation. This pattern supports corresponding research that shows the anesthesia providers' contaminated hands play a key role in bacterial transfer (Loftus et al., 2012). This a major concern for infection control because not all of the bacteria on surfaces and objects can be completely removed, so patients are at risk of being infected by the bacteria that is immediately present on surfaces and objects within the anesthesia environment (Stackhouse et al., 2011).

Anesthesiology, health technology, and healthcare in general will continue to grow in complexity, and as this occurs, it is crucial to reduce and ultimately eliminate the risk of infection in the OR. The anesthesia environment and the anesthesia provider play key roles in the transmission of bacteria during the perioperative care of a patient. If anesthesia providers can reduce the number of surfaces and objects they come in contact with in the anesthesia environment, there can be a potential reduction in risk of bacterial transmission to the patient. It would be ideal, sterility-wise, to have all HCI be touchless in the OR, but this is not possible with the current work practices of anesthesiology and it's fast-paced environment. Although completely touchless HCI is futuristic, the technology currently exists to facilitate a number of touchless interactions through gestural communication.

There is an opportunity to determine if gestural input technology makes sense as an intervention for anesthesia providers in the OR to improve bacterial transmission and sterility. In order to do so successfully, gestures should be elicited from users (Morris et al., 2010) and be suitable for the context and domain in which it is applied (Ardito et al., 2014; Nielsen et al., 2004; Wigdor & Wixon, 2011). There has been some research investigating gestural control for anesthesia providers in the OR (Jurewicz & Neyens, 2017; Jurewicz et al., 2018); however, several barriers, relative to the technology and the human-system, still exist to the adoption of gestural technology in anesthesia and it's extension to other healthcare applications.

Barriers to Adoption of Gestural Technology

Technical Barriers

There are several challenges to developing gestural systems which are currently slowing its growth in the HCI field and application in healthcare. One challenge inherent in gestural systems is the tradeoff associated with the number of gestures in the vocabulary and performance of the system (Wachs, Kölsch, Stern, & Edan, 2011). The training and software development become increasingly difficult as the gesture vocabulary set grows (Anderson & Bischof, 2013; Ardito et al., 2014). Furthermore, camera-based gestural systems perform continuous capture of either hand or body movement; therefore, as the expanse of possible gestures and gesture combinations grow, there is a segmentation issue in the capture of the gestures (Baudel & Beaudouin-Lafon, 1993; Pickering, Burnham, & Richardson, 2007). The gestural input system needs to be capable of segmenting the movements to understand which gesture has actually been performed, and since the capture is continuous, it may become difficult for the gestural system to differentiate distinct gestures, especially if a specific gesture for a function is complex (e.g., dynamic, rotating gestures). Along the same lines of hardware limitations is the issue of occlusion. The cameras of the gestural system rely on a clear visual of the hand and fingers, and if a person or an object occludes the camera, the gesture cannot be captured (Rautaray & Agrawal, 2015). Lastly, one of the biggest challenges from the technical side is ensuring that gestures are accurately recognized, and there has been considerable interest in the research community to develop methods which ensure a high recognition accuracy. One example is developing a deep neural network which learns particular features for gesture recognition from the raw data from the camera (Huang, Zhou, Li, & Li, 2015). The deep neural network recognition approach has been shown to increase recognition accuracy to about 99% (Huang et al., 2015). Other methods have shown equal success such as hidden markov models, support vector machines, Eigenspace-based methods, and dynamic programming (Pisharady & Saerbeck, 2015).

Human Factors Barriers

Aside from these more technical concerns, there are several human factors related issues of gestural systems relating to context, expertise, and how interaction behavior may change over time via increased exposure with the gestural system. It has been shown that gestures are highly individualized (Stern et al., 2008), and that this individualization, specifically the interpretation of a gesture, is dependent on an individual's culture and past experiences (Mauney, Howarth, Wirtanen, & Capra, 2010; Rautaray & Agrawal, 2015). The differing gesture interpretations make it difficult to create a gesture vocabulary that is intuitive for all users. Another challenge when developing a gesture system is considering that a gesture interpretation may differ from context to context (Ardito et al., 2014; Jacob & Wachs, 2014a; Jacob et al., 2013; Nielsen et al., 2004; Wigdor & Wixon, 2011). Gestural systems are not expected to succeed if general gesture vocabulary sets are used for HCI across multiple contexts (Ardito et al., 2014). Thus, gestural interfaces must consider the context in which it will be used (Ardito et al., 2014; Jacob & Wachs, 2014a; Jacob et al., 2013; Nielsen et al., 2004; Wigdor & Wixon, 2011) and incorporate new possibilities that the gestural interaction could bring to that context (Wigdor & Wixon, 2011). Some user-elicitation studies have revealed the issue of context sensitivity in that multiple functions are mapped to the same gesture (Jurewicz & Neyens, 2017; Jurewicz et al., 2018; Pereira, Wachs, Park, & Rempel, 2015). If there is overlap in gesture-function mappings then the recognition software needs to be aware of the context in which a gesture is used in order to complete the correct task.

All users of a gestural system, from novice to expert, should be able to use the system (Wigdor & Wixon, 2011); therefore, an additional challenge to gestural system design is ensuring that multiple user groups have an intuitive and natural experience. There has only been one study investigating the gesture behavior of both experts and novices within a specific context, and it was shown that domain expertise is influential when generating gesture-function mappings in a study investigating gestural control for anesthetic tasks in the OR (Jurewicz et al., 2018). Domain experts' gesture choices tended to be influenced more by physical components in the anesthesia environment whereas the gesture choices generated by domain novices did not show a relationship to the physical environment (Jurewicz et al., 2018). Domain novices additionally demonstrated longer reaction times in generating gestures potentially suggesting that there is a greater cognitive load for those gesture-function mappings (Jurewicz et al., 2018).

Since there is evidence that shows expertise is influential in gesture choice, it may be important to capture how experience with the gesture system is gained over time and how gesture responses may change as expertise grows or as a user becomes more familiar with the system. However, no studies have investigated the effect of exposure to the system over time on gesture behavior. Overall, there is minimal, if any, evidence addressing human factors challenges as a whole to gestural system design, and it remains unclear how to overcome these challenges in order to implement an intuitive, reliable, and natural gestural system. The current develop approaches to 3D gestural systems is to either take a technology-based approach or human-based approach (see Figure 2).

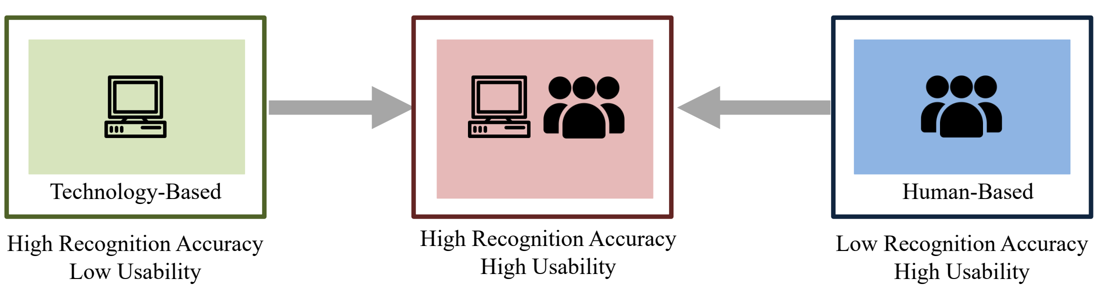


Figure 2. Illustration of the tradeoffs between recognition accuracy and usability in current gesture development techniques

However, there is a tradeoff between recognition accuracy and usability. Ideally, the systems would be both highly accurate and highly usable as shown in the middle of Figure 2. Thus, the purpose of this dissertation is to the further understand the arrow between the human-based approaches and move towards the goal of the middle box with highly accurate and highly usable 3D gestural systems.

Skills-, Rules-, and Knowledge-Based Behavior

Using the hands as an input modality in HCI is relatively new to the science of human factors, so it is important to understand how humans perform different tasks in information systems. The skills-, rules-, and knowledge- (SRK) based behavior taxonomy (see Figure 3), developed by Rasmussen (1983), may be utilized to explain how context, expertise, and exposure influence the human's behavior in a gestural system.

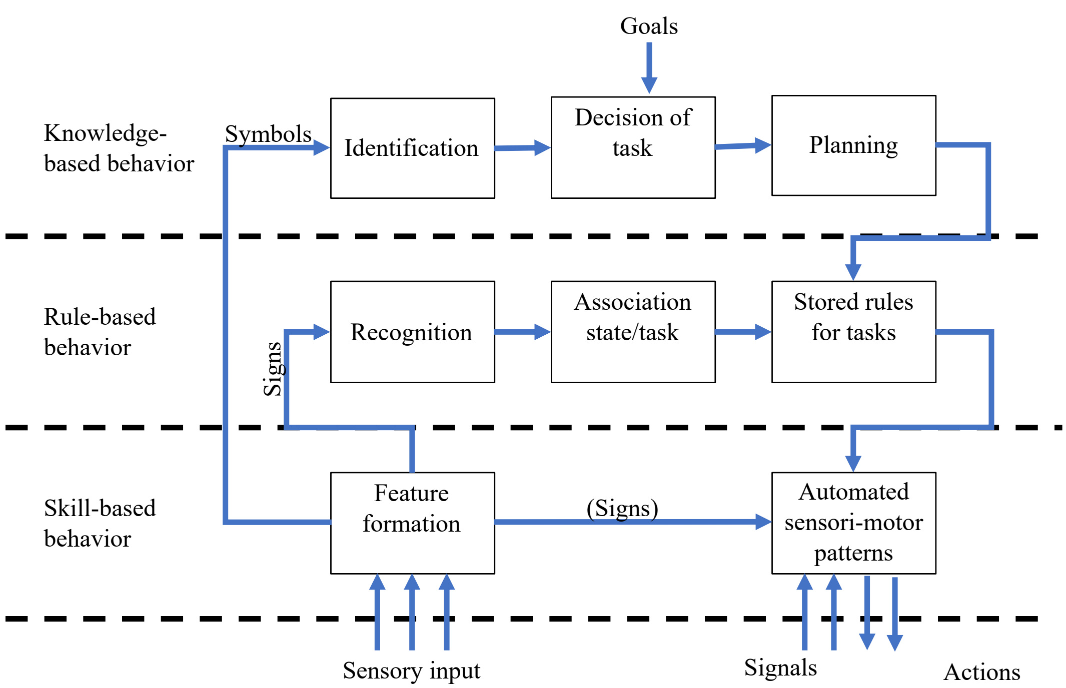


Figure 3. Skills Rules and Knowledge-Based Behavior framework. Figure image derived from (Rasmussen, 1983)

The SRK taxonomy was built in order to describe human performance in different task conditions, whether tasks be routine or unfamiliar (Rasmussen, 1983). Human behavior can be categorized into knowledge-based behavior, rule-based behavior, and skill-based behavior (Rasmussen, 1983). Knowledge-based behavior is most often practiced in unfamiliar situations and is commonly seen in novice users or in novice situations (Rasmussen, 1983; Vicente & Rasmussen, 1992). A user in knowledge-based behavior carefully develops a goal, analyzes the environment, and formulates a plan for acting on the system (Rasmussen, 1983). At the next level is rule-based behavior where actions are dominated by a stored rule that is driven by previous experiences or expectations of the system (Rasmussen, 1983). Rule-based behavior requires less analysis and conscious thought than knowledge-based behavior (Rasmussen, 1983). At the highest level of SRK is skill-based behavior where actions are executed without conscious effort (Rasmussen, 1983). Skill-based behavior is often practiced in very familiar situations where users react automatically to input from the environment (Rasmussen, 1983).

The discrepancy between the two ends of SRK lies in the applied cognitive effort. With knowledge-based behavior, a goal is explicitly formed and conscious effort is put towards planning how to achieve the goal, whereas a user in skill-based behavior, who has performed this action many times in previous experiences, will go straight from the sensory input to an intuitive action (Vicente & Rasmussen, 1992). Skill-based behavior is highly automated, and the body and environment work fluidly together (Rasmussen, 1983).

The SRK taxonomy was not developed to be explicitly context sensitive; however, context may influence how information is perceived in the environment and may influence the decision process and action chosen. In the end, the decision process and action chosen depends on the user's familiarity with the system and their prior beliefs, emphasizing the role of expertise. Users often must go complete the decision process and complete an action multiple times thus extending the SRK taxonomy from a single interaction to multiple interactions via exposure over time. Users do not stay in the same level of cognitive control throughout interaction with a system due to varying task demands, and users rather fluctuate between knowledge-, rule-, and skill-based behavior (Vicente & Rasmussen, 1992). Therefore, when designing a new HCI, it is suggested that the goal is to support all three levels in the SRK taxonomy and to avoid forcing a user to a certain level of cognitive control (Vicente & Rasmussen, 1992).

The SRK taxonomy is a means of understanding how human behavior may adapt as context, expertise, and exposure change over time, especially when investigating a new interaction method such as 3D gestural control. Understanding a user's gesture behavior is valuable to researchers and designers as this data can be applied in statistical models to predict how users may gesturally respond to a computer system based on specific individual factors. Designers can more completely understand the end user and implement intuitive gesture systems that are based on expected natural responses if gesture choice can be predicted accurately. If designers can anticipate and predict gesture choice for users, then the system can be designed to support multiple gestural interactions thus affording an intuitive and natural experience for multiple user groups.

Bayesian Framework

Several statistical techniques could be applicable to understanding and predicting gesture behavior. Bayesian statistical approaches may be most suitable for understanding novel input systems, such as integrating 3D gestural input into HCI. Gestures are difficult to learn (Hinckley et al., 2014), and there is little intuitive gestural agreement among users (Wobbrock et al., 2009). In other words, there is much uncertainty when it comes to gestural interaction from an observer's standpoint. A Bayesian analytical approach may be useful in such scenarios because Bayesian statistics provides a quantitative framework for modeling uncertainty. Bayesian statistical approaches are able to represent knowledge of the current state of the system and then integrate new information to understand a new or updated state of a system (Cowles, 2013).

Bayesian statistics is not currently as widely used in human factors research as frequentist approaches despite its several advantages. Frequentist approaches rely on many assumptions that may difficult to meet when attempting to predict a user's gesture behavior. In frequentist statistics, an experimenter tests whether an event (the hypothesis) occurs or not. Experiments are repeated under the same conditions until a stopping point, and it is calculated whether there is enough evidence in the data to state if a hypothesis is rejected or if it failed to be rejected. The key to frequentist statistics is the concept of repeating experiments over and over again and having a large enough sample size or "frequency" to come to make inferential statements, and it is possible that there are different outcomes depending on the stopping point of the experiment. The calculation of point estimates, confidence intervals, and p-values may differ depending on when an experiment is stopped. These calculations may additionally continue to differ as sample sizes change and as the experiment is repeated.

Bayesian statistics is different approach from frequentist statistics from modeling data, to the calculations, and to how the data is modeled conceptually. From the computational side, both Bayesian and frequentist approaches account for the likelihood of data. However, Bayesian statistics uses probabilities to quantify prior beliefs whereas frequentist statistics does not. Before collecting or gathering data, there is a belief about the outcome of interest and this belief is modelled in terms of a prior distribution. The prior may have different functional forms depending on the scenario. After data is observed, the belief is then updated through computation of the posterior distribution, and then the posterior distribution is used to make inferences about the data. Bayesian statistics can be analytically advantageous as it is capable of incorporating one's beliefs about the data. Analytically, this concept is grounded in Bayes' theorem which defines the relationship between prior and posterior probabilities in that the posterior probability of a model is proportional to the prior probability times the likelihood (graphically shown in Figure 4):

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

where:

$P(\theta)$ is the strength of the current belief of the system (i.e., the prior belief)

$P(D|\theta)$ is the likelihood of observing our result given the distribution of our belief

$P(D)$ is the new information

$P(\theta|D)$ is our new belief of the system after gaining the new information (i.e., posterior belief)

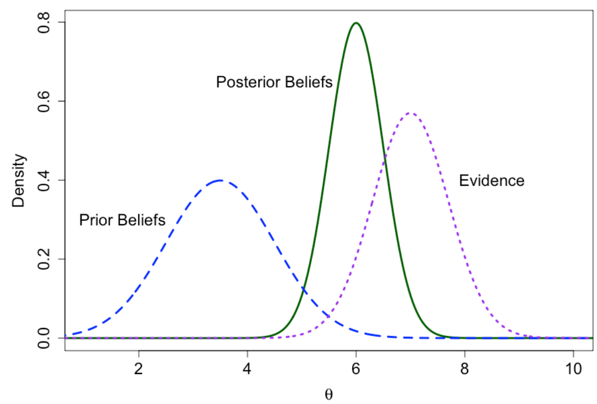


Figure 4. Graphical Representation of Bayes' Theorem

There are parallels between the Bayesian mindset of incorporating prior beliefs with data and how humans process information and make decisions through the SRK taxonomy. As new information is gained or as different users process information in either skill-, rule-, or knowledge-based behavior, a new belief of the system is formulated in order to make a decision on how to act upon that system. For example, consider a driver approaching a yellow and then a red traffic light. At a previous instance, the driver speeds through the yellow light so as to miss the red light but does not do so fast enough and crashes into a car travelling perpendicular to them. The driver develops a new belief in the system and next time they see a yellow light, they decide to slow down earlier or decide to perform more visual checks before trying to speed through the yellow light.

The same process of updating beliefs and its influence on decision making can be extended to how humans interact with computers, especially with novel input modalities such as 3D gestural displays. Consider any given person interacting with a computer with hand gestures for the first time. The gesture chosen can be represented by the user's belief in the system (i.e., it represents the prior and subsequently posterior beliefs after data has been gathered), and it is also the action decision in terms of SRK-based behavior. Given the context and expertise within the context, the intuitive gesture choice (i.e., the SRK-based behavior action or the Bayesian posterior inference) may update over time (i.e., the effect of exposure) as the user gains information about the system and the interactions. Thus, the 3D gestural interactions can be explained by the SRK taxonomy in how actions are chosen but also be translated analytically through the Bayesian mindset to model the decision making process and eventually make predictions on future use of the system based on prior data.

Research Objective and Questions

The objective of this research is to understand how expertise and exposure influence gestural behavior and to develop Bayesian statistical models that can accurately predict how users would choose intuitive gestures in anesthesia based on expertise and exposure. Expertise and exposure may influence gesture responses for individuals; however, there is limited to no work investigating how these factors influence intuitive gesture choice and how to use this information to predict intuitive gestures to be used in system design. Therefore, my dissertation research questions are:

> **Q1**- How does expertise influence how users gesturally respond to a computer system?

Q2 - How does exposure to the gestural system influence how users gesturally respond to a computer system?

Q3 - How accurately can intuitive gestures be statistically predicted?

Answering these research questions will provide a more complete understanding of the end users so that intuitive gesture systems that are based on expected user responses can be developed. If gestures can be predicted with a particular level of certainty, then displays and systems can be redesigned to support all levels of cognitive control as modelled by the SRK taxonomy. These questions will be answered in a series of experiments:

1. **Chapter 3 -** A preliminary repeated measures study investigating the effect of **domain expertise (Q1)** on gestural behavior within the context of anesthesia
2. **Chapter 4 -** A longitudinal study investigating the effect of **workload and exposure (Q2)** on gestural behavior for controlling a PowerPoint presentation
3. **Chapter 5 –** Development of Bayesian statistical models of expertise and exposure data from chapters 3 and 4 to accurately **predict gesture choice (Q3).**

CHAPTER 3: INVESTIGATING THE EFFECT OF DOMAIN EXPERTISE ON GESTURAL BEHAVIOR WITHIN THE CONTEXT OF ANESTHESIOLOGY

Introduction

The work in this chapter contributes to addressing the first research question of this dissertation: Q1 – How does expertise influence how users gesturally respond to a computer system? Both expert and novice users should be able to interact with the gestural system naturally (Wigdor & Wixon, 2011), so it is important to capture the user behavior of both domain experts and domain novices as domain novices may also be asked to interact with anesthesia equipment at various points in the device's life cycle. Environmental context must be considered when eliciting gestures from users (Ardito et al., 2014; Jacob & Wachs, 2014b; Jacob et al., 2013; Nielsen et al., 2004; Wigdor & Wixon, 2011), and complete novices should be included from the very beginning of developing natural user interfaces (Wigdor & Wixon, 2011). Thus, if a gesture vocabulary is to be developed for the application of anesthesia tasks and functions in the OR, gestures should be elicited from both expert and novice users, within a representative environmental context, focusing on similarities and differences of gesture behaviors across users.

A repeated-measures study was conducted with two cohorts: anesthesia providers (i.e., domain experts) (*N*=16) and students (i.e., domain novices) (*N*=30). Participants chose gestures for ten anesthetic functions across three blocks to determine intuitive gesture-function mappings, and reaction time was collected as a complementary measure for understanding the mappings. This work sought to compare the mappings of gestures to functions generated for domain experts and novices when exposed to the same OR anesthesia context. The work of this chapter was presented at the 2016 Human Factors and Ergonomics Society Annual Meeting (Jurewicz & Neyens, 2017) and is published in *Human Factors* (Jurewicz et al., 2018).

Methods

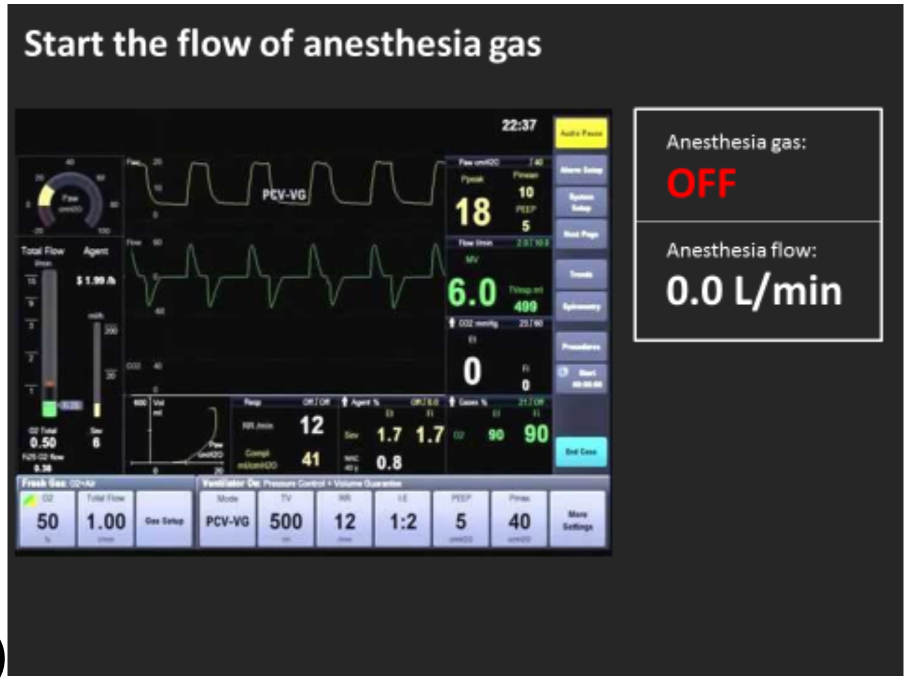
Nielsen Störring, Moeslund, & Granum (2004) created a human-centered procedure for developing intuitive and ergonomic gesture interfaces. As part of this procedure, gesture vocabularies can either be elicited from end users in a bottom-up or top-down fashion. The bottom-up approach presents functions and identifies matching gestures, and the top-down approach presents gestures and identifies a function mapping (Nielsen et al., 2004). The top-down approach is more suitable for testing a gesture vocabulary (Nielsen et al., 2004), so the bottom-up approach was used in this study since the goal is to generate gesture-function mappings,. In this approach, the function is shown to the user, and the user chooses a gesture that they believe maps to the function. The gesture that is most frequently performed across all users is mapped to a function as the most intuitive gesture. Therefore, this approach is a human-based, specifically one that reaches consensus, investigating two cohorts: the responses of undergraduate and graduate students and the responses of anesthesia providers. This research was approved by Clemson University IRB (IRB#: 2016-110) to study the responses of undergraduate and graduate students. This research was additionally approved by The Medical University of South Carolina IRB (IRB#: Pro00048787) to study the responses of anesthesia providers. The two cohorts were studied at separate times due to the availability of novice and expert participants, and the distance between the hospital and the university.

Participants

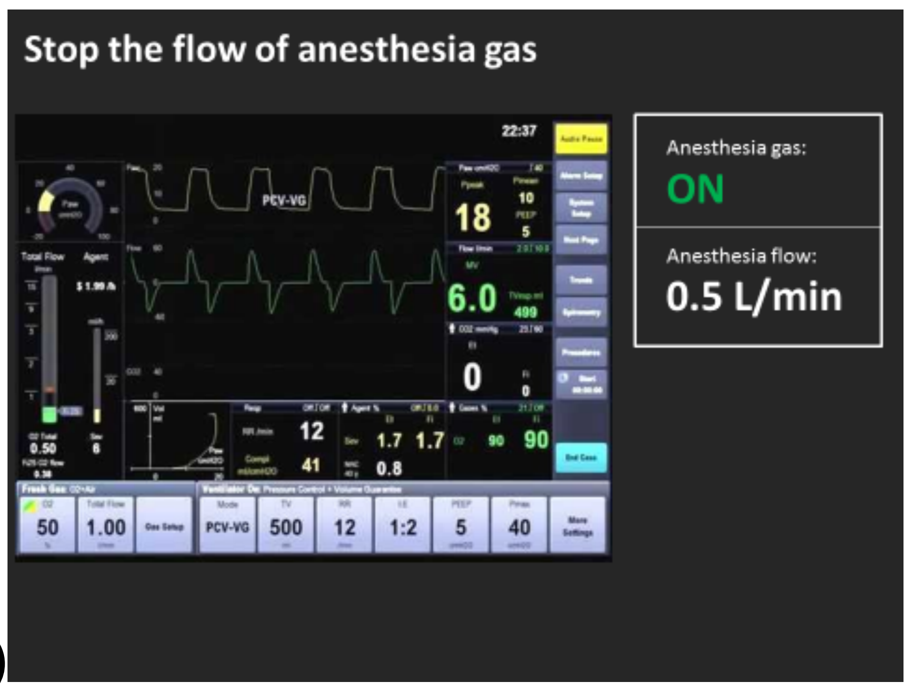
All participants needed to be able to move their fingers, wrists and arms without issue in their non-dominant hand and needed to be able to read, write, and speak in English. Participants were domain novices (*N*=30) and domain experts (*N*=16). The domain novices were undergraduate and graduate students, and the domain experts were anesthesia providers that included attending anesthesiologists, certified registered nurse anesthetists (CRNAs), and anesthesia residents.

Study Design

This study employed a repeated measures design where the functions (*N*=10) were repeated across three blocks. The functions are described in Figures 5, 6, 7, 8, and 9. Each block included all ten functions, and the presentation order of the functions was randomized within each block. The function displays were placed in a PowerPoint presentation according to a randomized order for each participant. The functions tested in the experiment were representative of typical tasks done by anesthesia providers in the OR and were selected after performing in-person and video observations in the OR (Betza et al., 2016). Some functions were generic examples used to elicit gestures for question answering (e.g., “Is the heart rate normal?”), and Function 9 was used for making choices among different options. Function 9 (See Figure 9a, “Select Heart Rate”) is not a task currently done by anesthesia providers but is a function that could be implemented as part of a 3D gestural system for anesthesia, as it is recommended to test new gestural functionalities when building gestural systems (Wigdor & Wixon, 2011).



a)



b)

Figure 5. Functions shown to participants in gesture elicitation experiment. (a) Function 1 - Start the flow of anesthesia gas. (b) Function 2 - Stop the flow of anesthesia gas

a)

b)

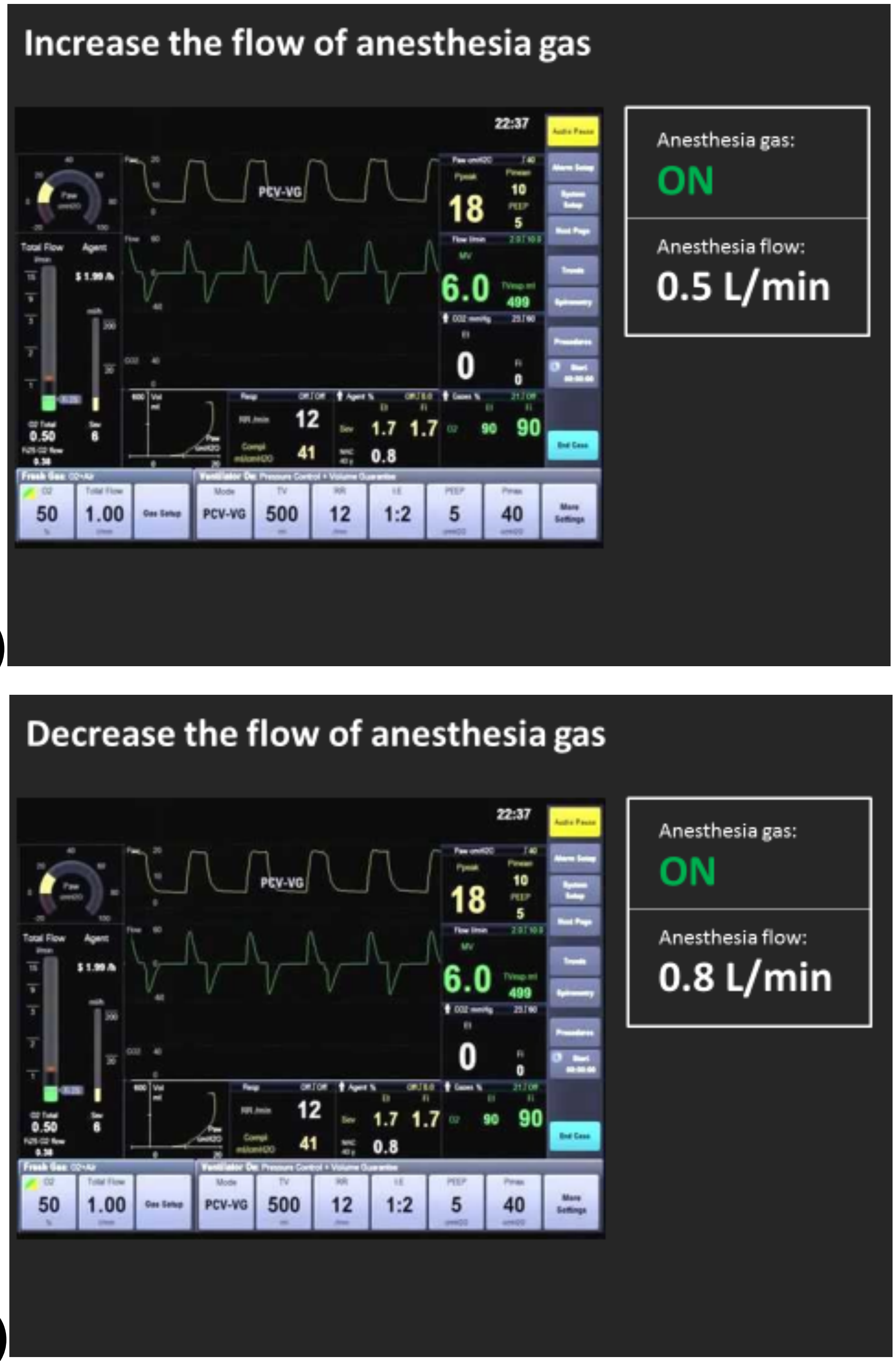
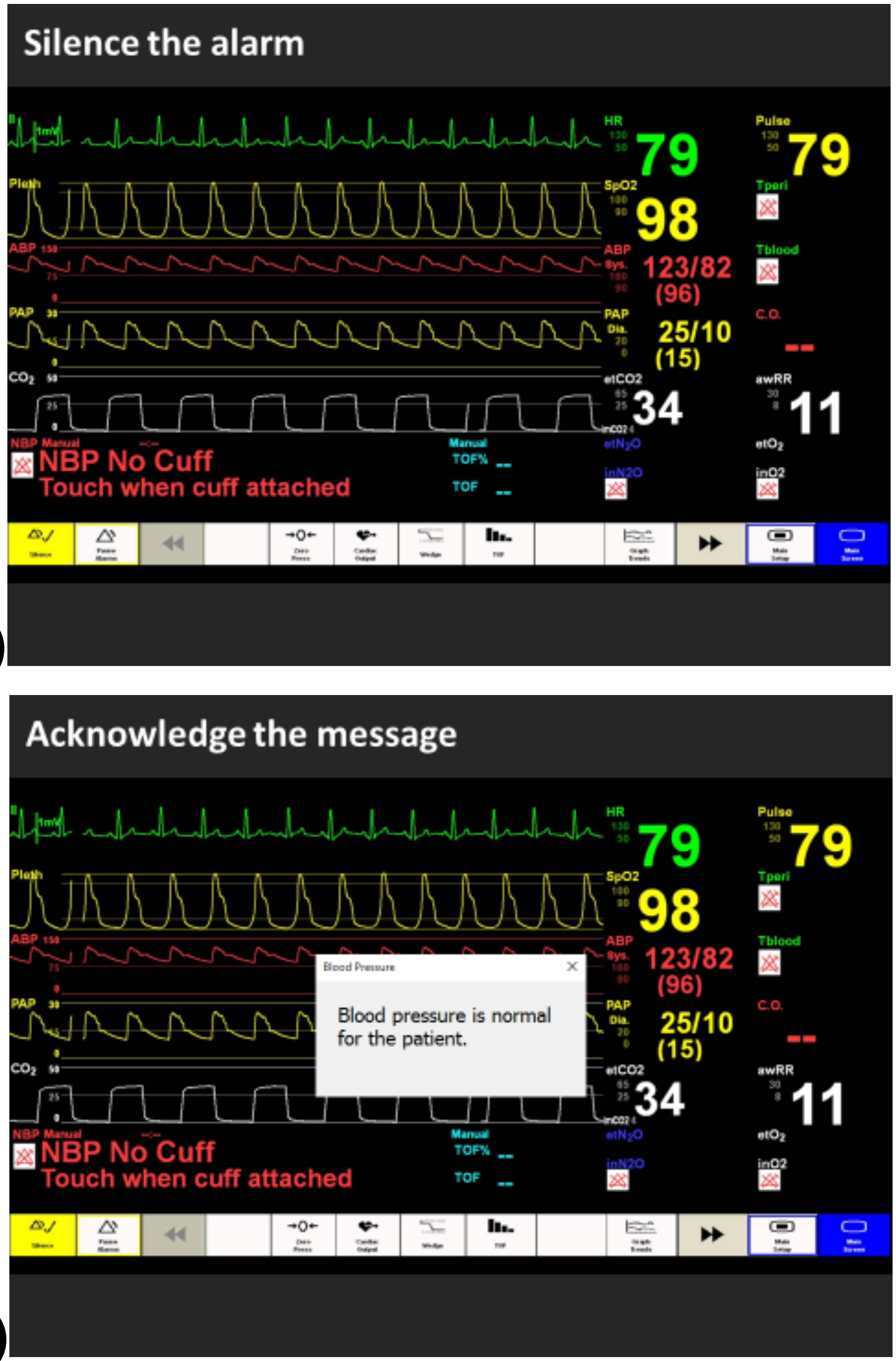
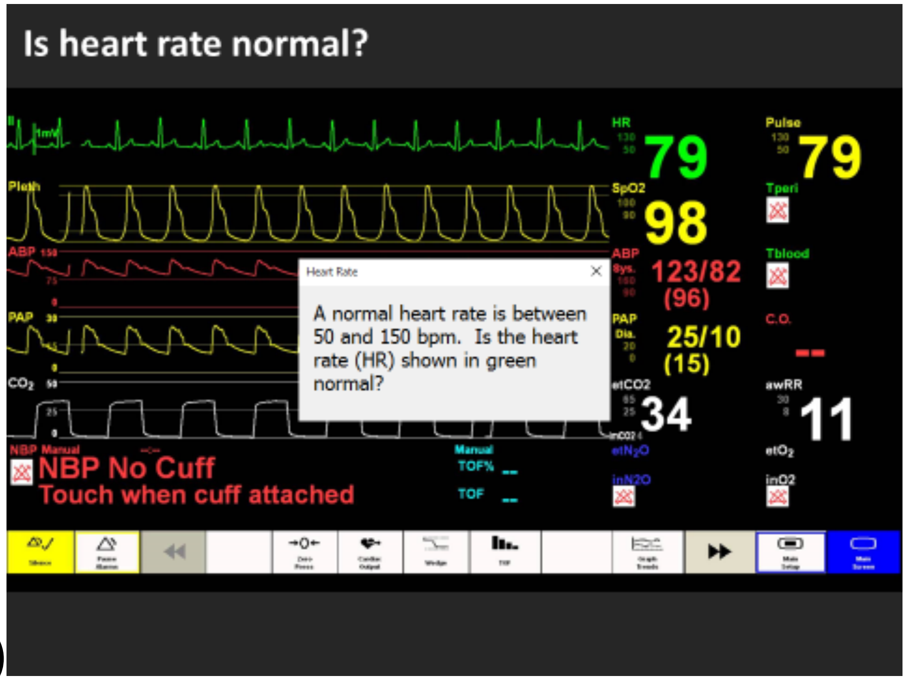


Figure 6. Functions shown to participants in gesture elicitation experiment. (a) Function 3 – Inc. the flow of anesthesia gas. (b) Function 4 - Dec the flow of anesthesia gas

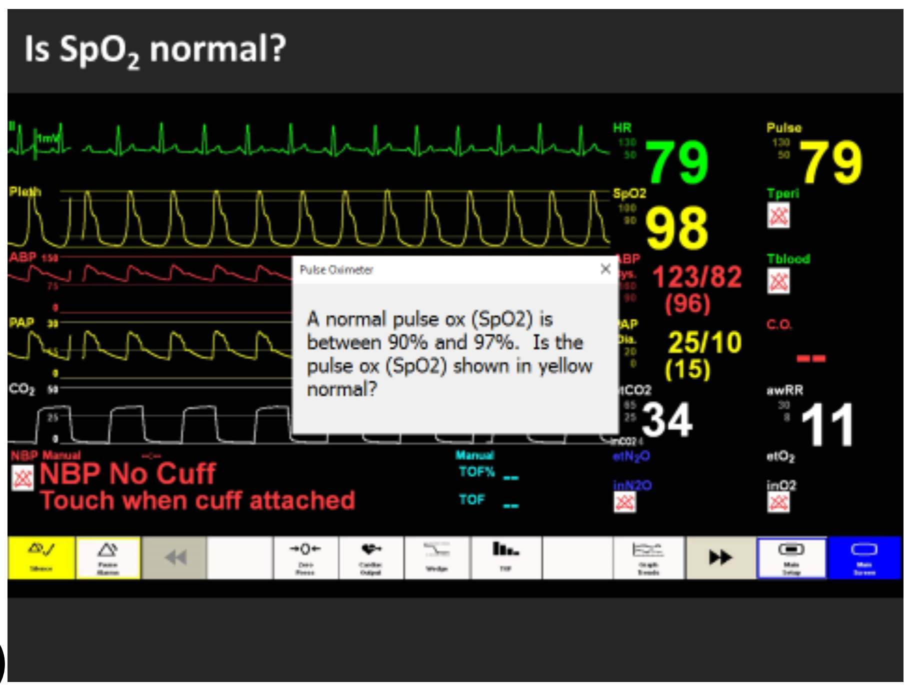


a) b)

Figure 7. Functions shown to participants in gesture elicitation experiment. (a) Function 3 - Silence the alarm (b) Function 6 - Acknowledge the message

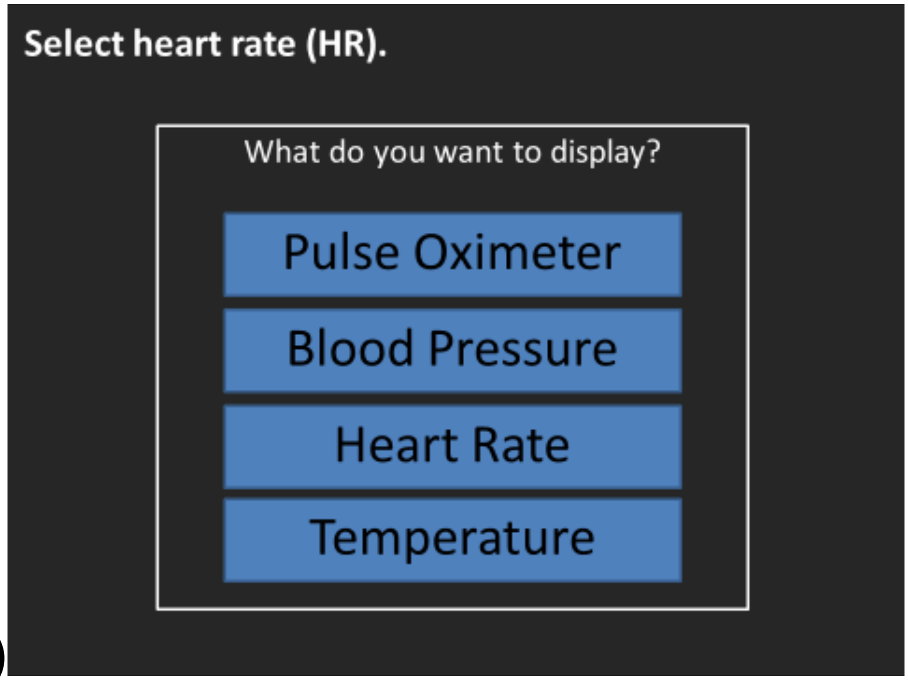


a)

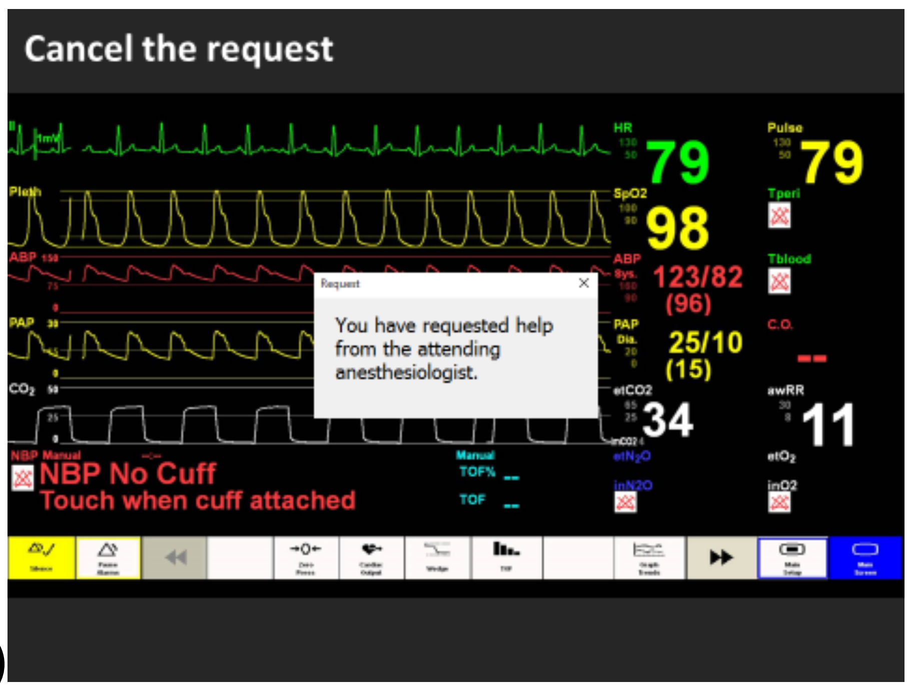


b)

Figure 8. Functions shown to participants in gesture elicitation experiment. (a) Function 7 – Is heart rate normal? (b) Function 8 – Is SpO2 normal?



a)



b)

Figure 9. Functions shown to participants in gesture elicitation experiment. (a) Function 9 - Select heart rate (b) Function 10 – Cancel the request

Equipment

The study of the two cohorts were completed at the same table with the same standard desktop computer with two monitors side by side (Figure 10). The study equipment (i.e., desk, two Dell 22-inch LED monitors, an Intel RealSense F200 Camera gestural camera, a PC running Windows 10, and medical gloves) was used at both locations with the position of the monitors and the gestural camera marked on the desk. Participants primarily interacted with the right monitor as this monitor presented the function displays and had the 3D camera attached on the top. A digital clock with the computer system time and depth-feedback of the 3D camera view were displayed on the left monitor. The setup of the computer and monitors did not differ between novices and experts; however, the study occurred in different rooms due to the participants being located at either the hospital or the university. The domain experts were in a conference room at the hospital that had additional tables, chairs, and a TV. The domain novices were at the university in an experimental room in a research lab without windows.

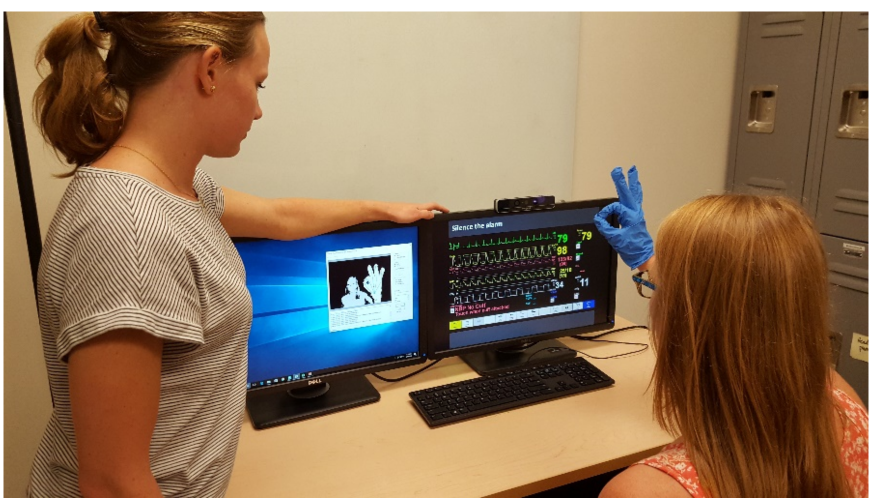


Figure 10. Experimental setup

The experimental session duplicated certain features of an anesthesia setting in the OR by sounding continuous and intermittent patient alarms and by having participants wear medical gloves. The World Health Organization (2009) recommends healthcare providers wear gloves when working with a patient, so wearing the gloves helped emulate anesthetic work. The alarms additionally helped to establish environmental context.

Procedure

The same study procedure was followed for both cohorts. Upon arrival, the informed consent process was completed and the participant filled out a demographics survey and the Complacency Potential Rating Scales (Singh, Molloy, & Parasuraman, 1993). There were two different demographic surveys created due to the differences in characteristics of the two cohorts. For example, anesthesia providers were not asked any questions about their major as these questions were not applicable. After completing the demographic survey, the participant familiarized themselves with the technology by practicing with the set of 14 gestures provided by the Intel RealSense SDK (Intel Corporation, 2016). Each gesture was performed 15 times according to Nielsen et al.'s (2004) approach for assessing the comfort of gestures in a user elicitation study.

The participant then completed the experimental task. A "Wizard of Oz" technique was used in the experimental session, which has shown to be valuable in gesture user-elicitation studies (Aigner et al., 2012; D. Freeman, Benko, Morris, & Wigdor, 2009; Höysniemi, Hämäläinen, & Turkki, 2004; Morris et al., 2010). In the

Wizard of Oz technique, the experimenter takes the place of an automatic system, interpreting inputs and controlling outputs. The manual control is done in order to evaluate functions and interfaces prior to investing in the technology required for automatic input and output. In the experiment, the system is perceived to be controlled by a participant's gestural input, but the experimenter manually progresses to the next function after a gesture is performed; therefore, this is not a complete Wizard of Oz study as the experimenter is physically in the room with the participant. Having the experimenter manually progress to the next function generates an effect-cause relationship between gesture and function that would be expected if the gestural system were actually implemented and working. The function display (the effect) was always presented first and then the participant would choose a gesture (the cause) that they believed initiated the function. Participants performed gestures of their choosing and whichever gesture was their "first guess" to complete the function.

Intuitive Gestures Measure

The intuitive gesture-function mappings were analyzed separately for the experts and novices in order to identify the differences between the two gesture sets. Videos of the participants' hands and fingers were recorded and analyzed to determine which gestures were performed for each function. A list of potential gestures performed was built by the research team to aid in the gesture analysis. The gesture list included the name and definition of all gestures used in the practice session, gestures from other studies, and commonly known cultural gestures. The list of potential gestures was created to provide standardization in gesture classification among the researchers. All videos were analyzed by three researchers separately, and gestures were classified according to the best-fit definition in the gesture list. Any discrepancies in gesture classification were discussed until the three researchers agreed on which gesture was chosen by the participant.

Incomplete gestures were removed from analysis. According to Nielsen et al.'s (2004) approach, the intuitive gesture for a function is the gesture that is most frequently chosen across a group. The gesture responses for each function were compiled in a table, and the gesture response that was performed most frequently across the experimental group was chosen as the intuitive gesture-function mapping.

Reaction Time Measure

The reaction time from presentation of the function display to completion of a gesture was recorded for every gesture-function pair by a Visual Basic program embedded in PowerPoint. This data was collected to complement the analysis of the intuitive gesture-function mappings because it has been shown that shorter reaction times are associated with higher convergence of gestures performed for a function (Pereira et al., 2015). The reaction time data was combined into one analysis for the experts and novices. A mixed linear regression model with participant ID as the random effect was used to identify gesture-function mappings that exhibited longer reaction times. A mixed linear regression model was used in order to account for both fixed and random effects. The fixed effects in the model were handedness, video game experience, virtual reality experience, the functions, and participant type. Interaction effects between function and participant type were included in the model to identify differences between experts and novices. Only reaction times from the first block were analyzed in order to separate the first instance the participant was exposed to a function and to avoid any issues in the statistical model related to learning effects that could be present in the other blocks. The equation below shows the mixed effects linear regression model in matrix notation.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon}$$

where:

$\boldsymbol{y}$ is an *N x 1* column vector of the response variable

$\boldsymbol{X}$ is an *N x p* matrix of *p* predictor variables

$\boldsymbol{\beta}$ is a *p x 1* column vector of the regression coefficients of the fixed effects

$\boldsymbol{Z}$ is an *N x q* matrix of *q* random effects

$\boldsymbol{\gamma}$ is a *q x 1 column* vector of the random effects

$\boldsymbol{\varepsilon}$ is an *N x 1* column vector of the residuals

R version 3.2.2 was used for all data analysis; the *lmer* function of the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) was used to build the mixed linear regression model, and the ggplot2 package (Wickham, 2009) was used to plot the data. Before fitting this mixed effects model, an ANOVA was performed to compare two linear models: a linear model with a fixed intercept plus the random effect and a null model with only the fixed intercept. If the P-value is <0.001, then the mixed model was preferred over the null model. Insignificant variables were stepwise deleted to obtain the final model. Diagnostic tests were performed to ensure the assumptions for the linear model is met: linearity, homescedacity, normality, independence, and no multicollinearity issues. To identify any multicollinearity issues, variance inflation factors (VIFs) were calculated and any predictor variables with a VIF >5 were removed from the model. Any influential points were also removed from the data set by calculating Cook's distance. Cook's distance is a measure for one unit's influence on parameter estimates (Cook, 1977). The formula for calculating Cook's distance is shown below:

$$\mathrm{D}_i = \frac{{e_i}^2}{s^2 p}\left[\frac{h_i}{(1-h_i)^2}\right]$$

where:

D_i is Cook's distance for the *i*th observation

e_i is the residual for the *i*th observation

s^2 is the mean squared error of the regression model

h_i is the leverage of the *i*th observation

For mixed models, a point is regarded as influential if the respective Cook's Distance value exceeds the cut off value of (Van der Meer, Te Grotenhuis, & Pelzer, 2010):

$$4/n$$

where *n* refers to the number of groups of the grouping variable.

The mixed effects linear regression model can only determine if functions are associated with response times compared to one reference function, so in order to compute differences in response times for each pair of functions, Tukey contrasts were calculated to make the pairwise comparisons. R version 3.3.2 was used to do the analysis and used the *skewness* function of the e1071 package (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2015), the *lmer* function of the lme4 package (Bates et al., 2014), the *glht* function of the multcomp package (Hothorn, Bretz, & Westfall, 2008), and the *cooks.distance* function of the influence.ME package (Nieuwenhuis, te Grotenhuis, & Pelzer, 2012).

Results

The characteristics of the participants for both experiments are shown in Table 1. Some data were not collected in Experiment 2 that were collected in Experiment 1 to ensure that subjects could not be identifiable with the smaller sample size in Experiment 2 thus this data is absent from Table 1. The mean response time across all blocks for Experiment 1 was 4.77 seconds (SD=2.93 seconds), and the mean response time across all blocks for Experiment 2 was 4.31 seconds (SD=2.29 seconds). In Experiment 1, there could have been a total of 900 gestures taken into account for the analysis, but due to a lack of complete gestures in participants for some functions, only 849 gestures were analyzed. In Experiment 2, a total of 438 out of the 460 possible gestures were analyzed after removing responses that were incomplete. Overall, 42 unique gestures were performed in Experiment 1 and 27 unique gestures were performed in Experiment 2.

Table 1. Participant characteristics for novices (Experiment 1) and experts (Experiment 2)

Variable Name	Experiment 1 *N (%)*	Experiment 2 *N (%)*
Age	M=21.80, SD=2.23	-
Gender		
Male	15 (50)	-
Female	15 (50)	-
Handedness		
Right	26 (86.7)	14 (87.5)
Left	3 (10)	2 (12.5)
Ambidextrous	1 (3.3)	-
Education, highest degree obtained		
High School/GED	16 (53.3)	-
Bachelors	11 (36.7)	-
Masters	3 (10)	-
Education, area of study		
Science or Engineering	19 (63.3)	-
Not Science or Engineering	11 (26.7)	-
Video Game Use		
Yes	15 (50)	5 (31.3)
No	15 (50)	11 (68.7)
Virtual Reality Gaming Experience		
Yes	14 (46.7)	4 (25.0)
No	16 (43.3)	12 (75.0)

Intuitive Gesture Mappings

Bar graphs for the intuitive mappings are shown in Appendix A for the novices and Appendix B for the experts. A summary of the intuitive gesture-function mappings are shown in Table 2. Pictorial representations of the gestures that were intuitively mapped are shown in Figure 11. There were several functions with different intuitive mappings between the novices and the experts. Functions 1-5 were associated with different gestures for the novices and the experts. Functions 6-10 resulted in the same gestures for both the novices and the experts. Functions 6, 7, and 8 were all mapped to the "Thumbs up" gesture for both cohorts.

Table 2. Intuitive gesture-function mappings for novices and experts

Function [#]	Intuitive Gesture Mapped	
	Novices	**Experts**
Start the flow [1]*	Thumbs up	Rotate right
Stop the flow [2]*	Five up	Rotate left
Inc. the flow [3]*	Swipe hand up	Rotate right
Dec. the flow [4]*	Swipe hand down	Rotate left
Silence alarm [5]*	Swipe hand left	Push hand
Ack. the message [6]	Thumbs up	Thumbs up
Heart rate normal? [7]	Thumbs up	Thumbs up
Pulse ox normal? [8]	Thumbs up	Thumbs up
Select heart rate [9]	Push fingers	Push fingers
Cancel the message	Swipe hand left	Swipe hand left

Note: * indicates dissimilar mappings between the novices and experts

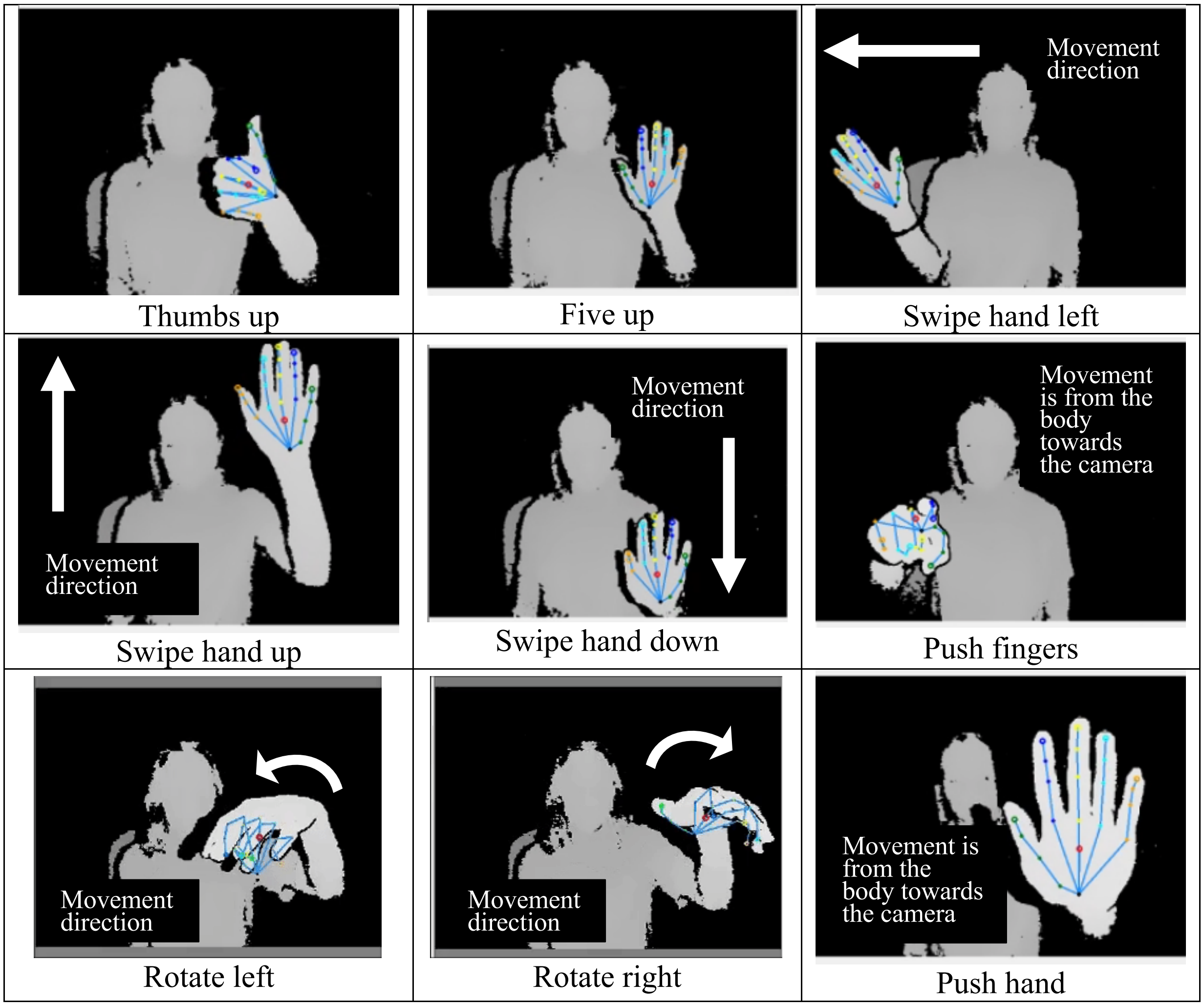


Figure 11. Pictorial representation of intuitive gestures mapped

Reaction Times

The mean reaction time for block 1 data for novices was 5.90 seconds (*SD*=3.66 seconds), and the mean reaction time for block 1 data for experts was 4.34 seconds (*SD*=2.28 seconds). Figure 12 shows the raw data of the reaction times for both groups.

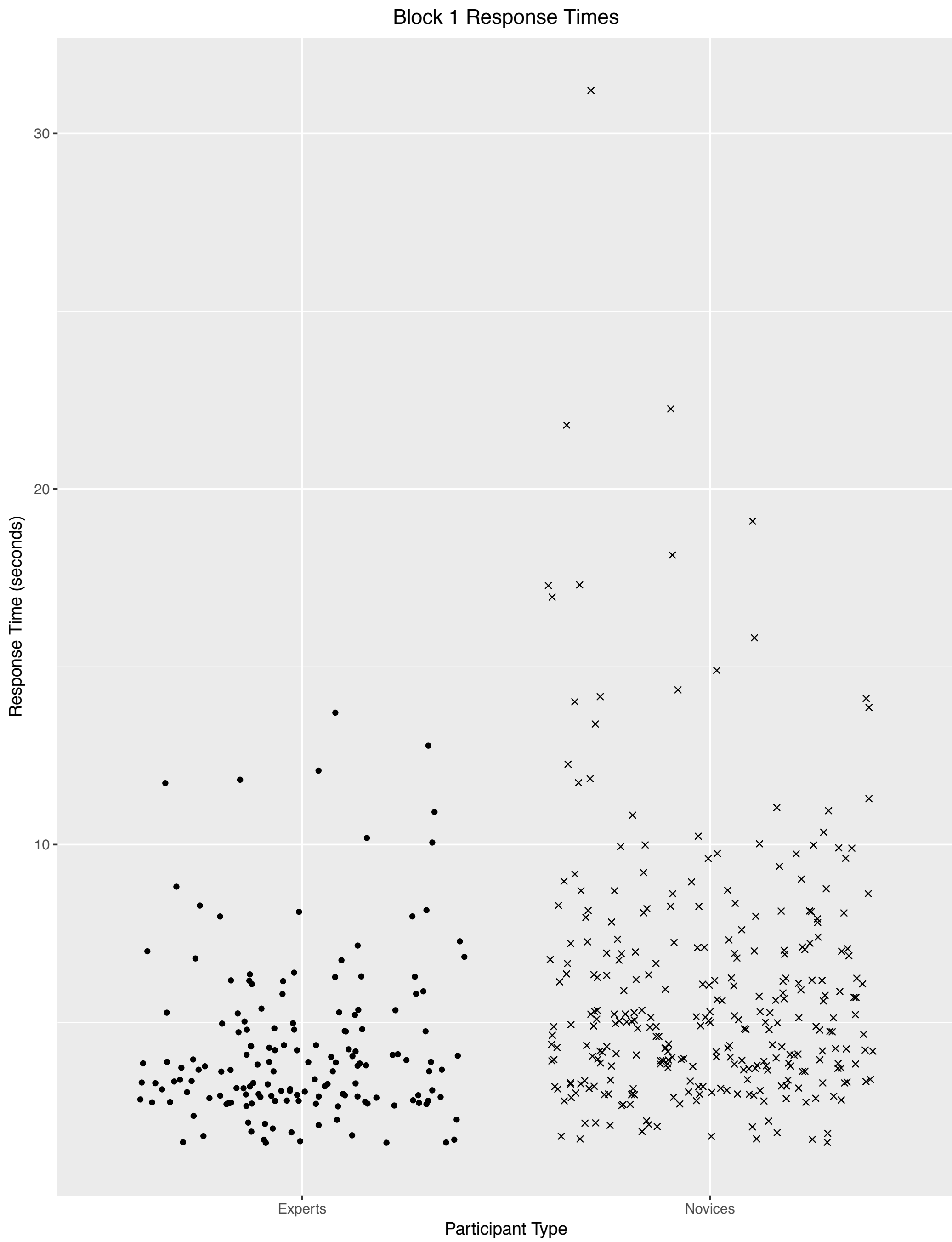


Figure 12. Jitter plot of response times for all participants in the first experimental block

Several assumptions needed to be met before moving forward with the mixed model regression analysis, including normality, testing random effects vs. fixed effects model, linearity, independence, homoscedasticity, and no multicollinearity issues. The skew of response times exhibited a positive skew with a value of 2.52, so the data was transformed by taking the natural logarithm of response times. After the transformation, the skew was 0.47. The ANOVA of the model comparison showed that the linear regression model with "Participant ID" as the random effect performed significantly better than the regression model with only the fixed intercept ($P<0.0001$); therefore, the mixed model was used for the analysis. With a mixed model, the within-subjects variability is removed as each participant is treated as a random effect; therefore, the assumption of independence of the data is met. The VIF values of this model were calculated and all VIF values were less than 5 indicating that there were no severe multicollinearity issues. There were no influential points in the data set as all of the calculated Cook's distances were below the cutoff value. The cutoff value for this dataset was $4/n=4/30=0.133$.

A summary of the final mixed linear regression models is shown in Table 3. Handedness, video game experience, and virtual reality experience were not significantly associated with longer reaction times and were thus stepwise deleted from the model. For variables with interactions, only the interaction terms are evaluated and main effects are not discussed. There were significant interactions between the user groups and some of the specific functions. Novices had significantly longer reaction times than experts for functions 7 through 10. These functions included: "Is heart rate normal?" ($p=0.007$), "Is

Pulse ox normal?” ($p<0.001$), “Select heart rate” ($p<0.001$), and “Cancel the message” ($p=0.005$).

Table 3. Results of mixed effects regression model for expertise study

	β	*SE*	*t*	$P >\| t \|$
(Intercept)	1.54	0.12	13.26	<0.001*
Stop the flow [2]	0.07	0.14	0.52	0.603
Inc. the flow [3]	-0.07	0.14	-0.48	0.634
Dec. the flow [4]	-0.08	0.14	-0.59	0.555
Silence alarm [5]	-0.32	0.14	-2.29	0.022*
Ack. the message [6]	-0.24	0.14	-1.73	0.084
Heart rate normal? [7]	-0.38	0.14	-2.76	0.006*
Pulse ox normal? [8]	-0.34	0.14	-2.45	0.015*
Select heart rate [9]	-0.25	0.14	-1.80	0.073
Cancel the message [10]	-0.51	0.14	-3.72	<0.001*
Novice	-0.01	0.17	-0.06	0.954
Stop the flow [2] x Novice	-0.31	0.17	-1.81	0.071
Inc. the flow [3] x Novice	-0.09	0.17	-0.51	0.611
Dec. the flow [4] x Novice	-0.11	0.17	-0.66	0.508
Silence alarm [5] x Novice	0.27	0.17	1.60	0.111
Ack. the message [6] x Novice	0.19	0.17	1.08	0.280
Heart rate normal? [7] x Novice	0.46	0.17	2.70	0.007*
Pulse ox normal? [8] x Novice	0.61	0.17	3.56	<0.001*
Select heart rate [9] x Novice	0.59	0.17	3.43	<0.001*
Cancel the message [10] x Novice	0.48	0.17	2.81	0.005*

Note: * p<0.05.

Discussion

The objective of this study was to evaluate the differences in intuitive gesture-function mappings between novices and experts. The students generated 40 unique gestures, and the anesthesia providers generated 27 unique gestures. The context did not change between anesthesia providers and students, and all participants were exposed to the same functions and displays, yet the gesture-function mapping sets differed between the anesthesia providers and the students. Five of the functions mapped to different gestures, and five mapped to the same gesture. The main finding of this study is that experts and novices differ in terms of intuitiveness of gestures thus emphasizing the need for domain expertise in the creation of a gesture vocabulary. Furthermore, the novice user group had significantly longer reaction times for four functions compared to the expert group.

There are characteristics of both the similar and different mappings that reveal insight into the gesture behavior of novices and anesthesia providers. For the set of functions that had different mappings (see functions 1-5 in Table 2), the anesthesia providers showed associations between the OR's physical environment and the gesture-function mapping. Specifically with the functions related to manipulating anesthesia gas, there were rotational gestures, similar to how anesthesia providers currently perform this task in the OR (Betza et al., 2016). Similarly, the "push hand" gesture of "Silence the alarm" is related to the physical interaction with the computers and monitors in the OR (Betza et al., 2016). Thus, the anesthesia providers' gesture mappings of these functions seem to show a strong contextual relationship to the physical environment. On the other hand, the gestures mappings which were the same (see functions 6-10 in Table 2) do not show this same level of association to the physical environment. The functions which had the same gesture mappings are more or less general human-computer interaction tasks such as cancelling, selecting, yes/no answers, and acknowledgement.

The differences in intuitive gesture-function mappings have design implications that should be considered when developing an intuitive context-specific gesture system. The anesthesia provider group exhibited a degree of domain expertise and contextual knowledge that was not inherent within the student group. Because of their expertise in the anesthesia domain and OR, the anesthesia providers chose gestures that were related to the anesthetic tasks as well as the physical and technological components in the anesthesia environment, such as rotational knobs and buttons. Conversely, the students demonstrated very few rotating gestures when the same contextual interface was presented to them. This suggests that in addition to context being important (Ardito et al., 2014; Jacob & Wachs, 2014b; Jacob et al., 2013; Nielsen et al., 2004; Wigdor & Wixon, 2011), domain expertise is also meaningful when creating the gestural vocabularies. However, the fact that both novices and experts chose similar gestures for half of the functions suggests that some functions may not necessarily depend on domain expertise. For gesture-function mappings that were the same across both user groups, potentially a more general population could be used to map gestures to functions, but a general population could not solely be used for all functions because of the gesture-function mappings which were different. The domain expertise of the anesthesia provider group generated about half as many gestures compared to the student group. Having a narrower set of gestures reveals some homogeneity within the anesthesia providers and may indicate convergence in gesture mapping agreements as a user group.

Additionally, the differences in reaction times between novices and experts for some of the functions further support our main finding that there is a need to consider domain expertise when building an intuitive gestural system. Longer reaction times may indicate that participants have difficulty generating a gesture-function mapping as previous studies have used reaction times as indicator for cognitive load (Horsky, Kaufman, Oppenheim, & Patel, 2003). This set of functions (7-10) all included language specifically related to the medical field (e.g., heart rate, pulse oximeter, attending anesthesiologist), and the lack of clinical knowledge in the novice group may have provoked longer reaction times among these functions. The longer reaction times may have also been due to a difficulty of generating a gesture-function mapping as reaction times may be used to indicate cognitive load (Horsky et al., 2003).

There are some limitations associated with this study. We were able to recruit 30 novices to do the study and only 16 experts, and this difference in sample size may have impacted the results. Specifically, the larger number of unique gestures generated in the novice cohort could be due to the larger sample of novices in the study. Furthermore, allowing participants to choose their own gestures for functions may have contributed to greater use of the same gesture for different functions; however, this was the most appropriate way to capture what gestures were intuitive to users by having participants perform their “first guess.” As part of our methodology the familiarization training with the technology may have influenced gestures chosen during the experiment. However, there was a large number of unique gestures recorded among the students (40) and among the anesthesia providers (27) and only 14 gestures were practiced as part of the familiarization training. Future research should evaluate how different practice gestures impact participant-derived gestures. Since this study showed that domain expertise is influential to gesture behavior, the next steps are to investigate how gesture behavior may change over time. The following chapters review data from a longitudinal gesture elicitation study that additionally investigated workload for Clemson-affiliated individuals controlling a PowerPoint presentation and show a Bayesian analysis of all the data from this dissertation to show how intuitive gesture choice can accurately be predicted.

CHAPTER 4: A STUDY INVESTIGATING THE EFFECT OF EXPOSURE AND WORKLOAD ON GESTURE BEHAVIOR IN A GENERAL HCI CONTEXT

Introduction

The work in this chapter contributes to addressing the second research question of this dissertation: Q2 – How does exposure influence how users gesturally respond to a computer system? This chapter additionally includes an investigation of workload on gesture behavior. There is little known about how gesture behavior may change over time or how gesture behavior may change when exposed to a high or low workload situation. This study investigated the gesture behavior of individuals giving a PowerPoint presentation on the history of the University. For this preliminary work, an anesthesia context was not used as none of the participants had domain expertise on anesthesia; however, all participants were affiliated with the University and using Microsoft PowerPoint. A mixed design study was completed with 40 participants for controlling a PowerPoint presentation about Clemson University. Participants were either assigned to a high workload scenario or a low workload scenario, and workload was counterbalanced across all participants. All participants chose gestures for all nine functions across three days and each day incorporated three blocks of functions.

Methods

This study incorporated Nielsen et al.'s (2004) human-centered approach for eliciting gestures, specifically for controlling a PowerPoint presentation similar to the approach used in Chapter 3. This research was approved by Clemson University IRB (IRB#: 2019-111).

Participants

All participants needed to be able to move their fingers, wrists and arms without issue in their hands and needed to be able to read, write, and speak in English. Participants were recruited at Clemson University and consisted of undergraduate students, graduate students, faculty, or staff. All participants were familiar with PowerPoint and the content of the presentation about Clemson University thus can be considered domain experts.

Study Design

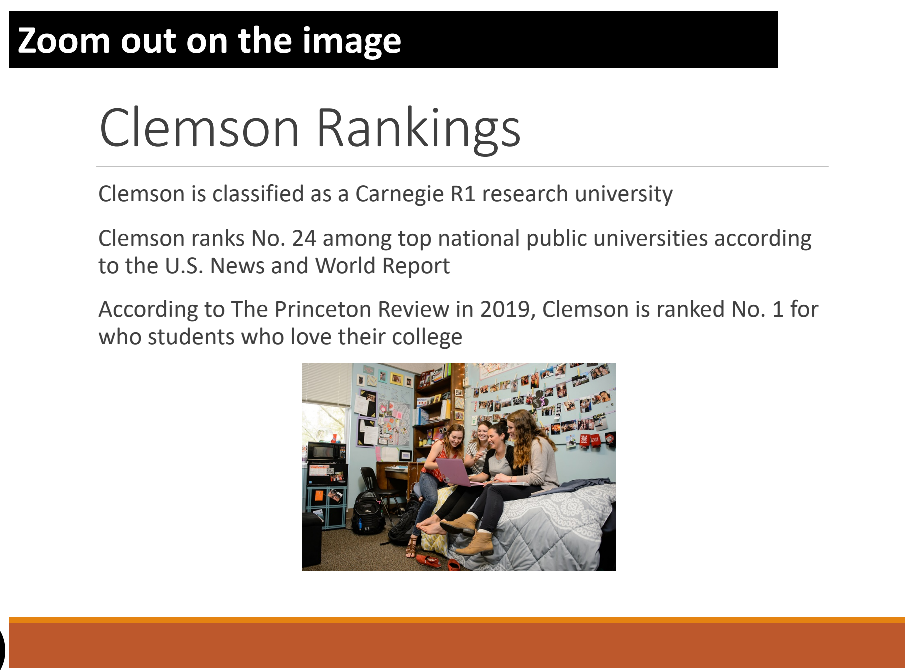
This study used a mixed design to investigate differences in workload and exposure. Participants were assigned to either a high or low workload condition and came to the lab three separate times to complete the study. The participants in the low workload condition only had to perform gestures for the functions similar to previous work (Jurewicz & Neyens, 2017; Jurewicz et al., 2018) and as described in Chapter 3. The participants in the high workload condition had to perform an improvised speech on a general topic in addition to performing gestures. The speech topics differed across the three experimental sessions: they did a speech on themselves in the first session (e.g., where they are from, what are their hobbies), a speech on their daily schedule in the second session (e.g., what classes they are taking or discussing their work), and a speech on their weekend plans in the third session. The functions (N=9) were repeated across three blocks in each of the session, resulting in 81 gestures for each participant. The functions are described in Figures 13, 14, 15, 16, and 17.The presentation order of the functions was randomized in every block. The function displays were placed in a PowerPoint presentation and randomized in every block.

a)

b)

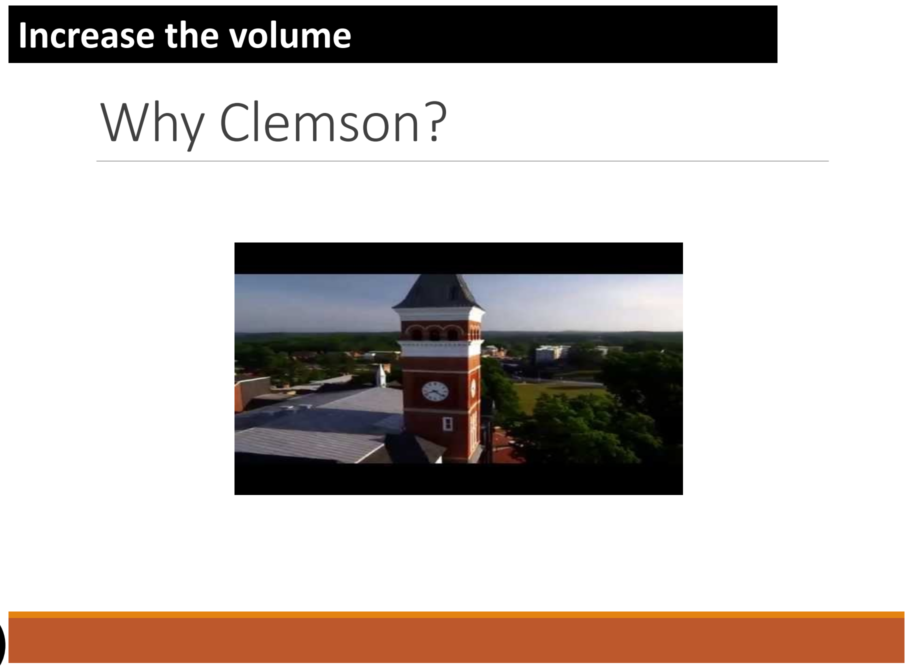
Figure 13. Functions shown to participants in gesture elicitation experiment. (a) Function 1 - Go to the next slide (b) Function 2 – Go to the previous slide

a)

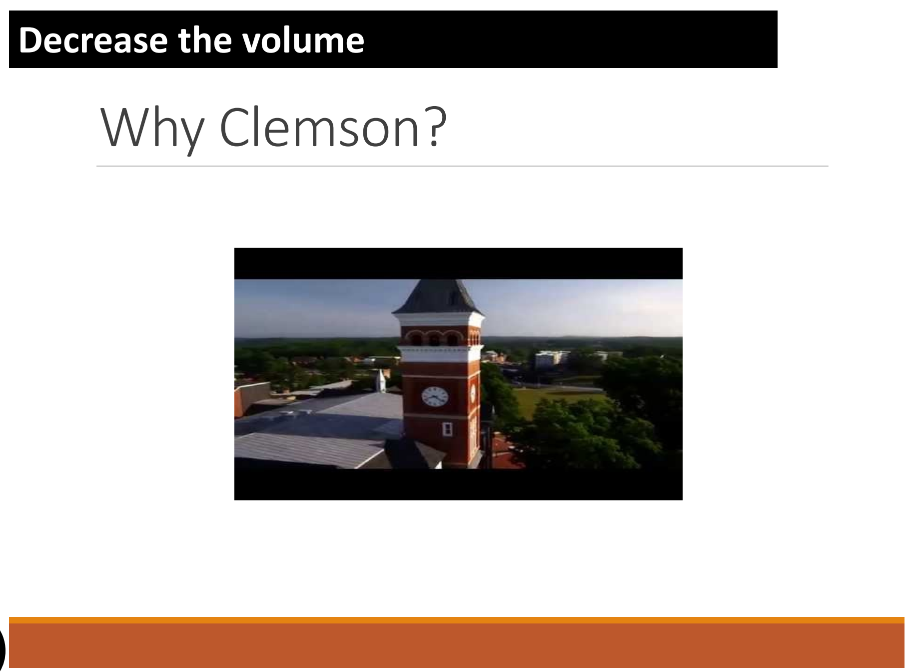


b)

Figure 14. Functions shown to participants in gesture elicitation experiment. (a) Function 1 - Zoom in on the image (b) Function 4 – Zoom out on the image

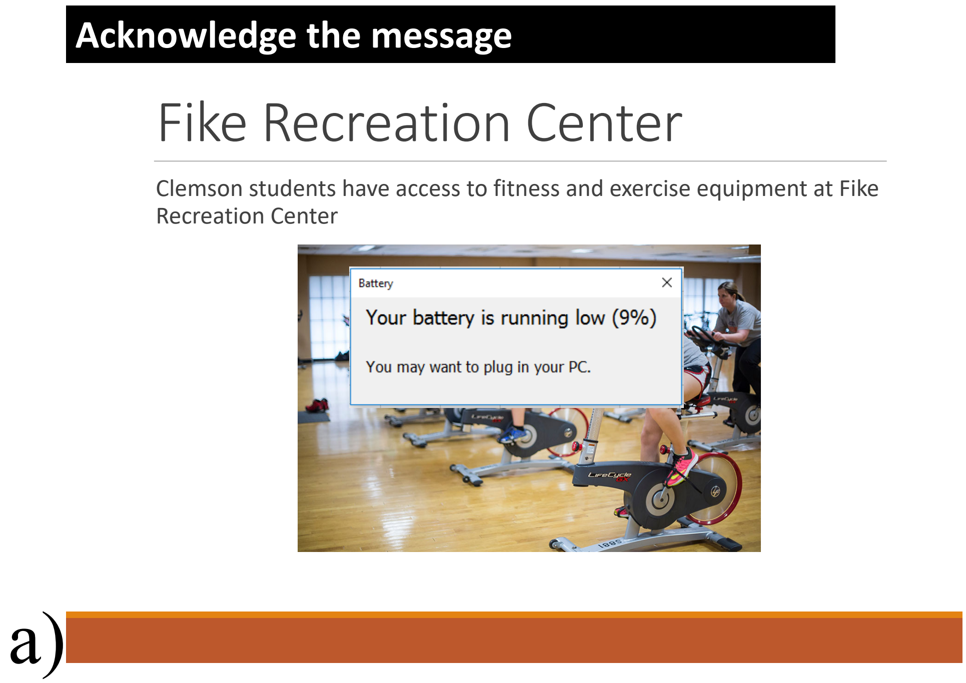


a)



b)

Figure 15. Functions shown to participants in gesture elicitation experiment. (a) Function 5 - Increase the volume (b) Function 6 – Decrease the volume



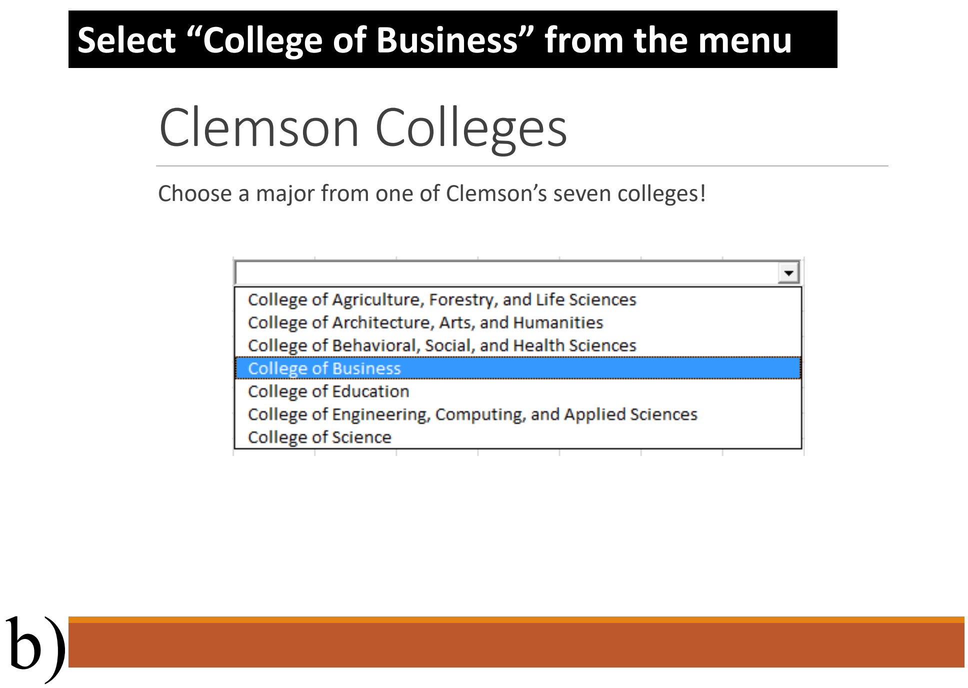


Figure 16. Functions shown to participants in gesture elicitation experiment. (a) Function 7 - Acknowledge the message (b) Function 8 – Select "College of Business" from the menu

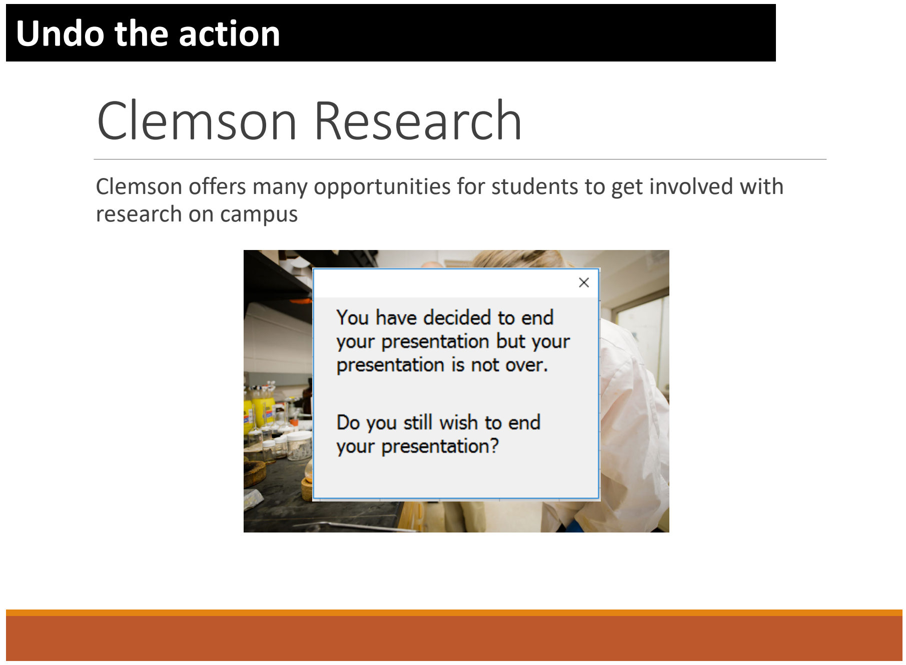


Figure 17. Functions shown to participants in gesture elicitation experiment. (a) Function 9 - Undo the action

Equipment

The experiment was conducted at a standing desk that resembled a podium (see Figure 18). There were two Dell 22-inch LED monitors, an Intel RealSense F200 Camera gestural camera, and a PC running Windows 10. Participants primarily interacted with the right monitor as this monitor presented the function displays and had the 3D camera attached on the top. A digital clock with the computer system time and depth-feedback of the 3D camera view were displayed on the left monitor.

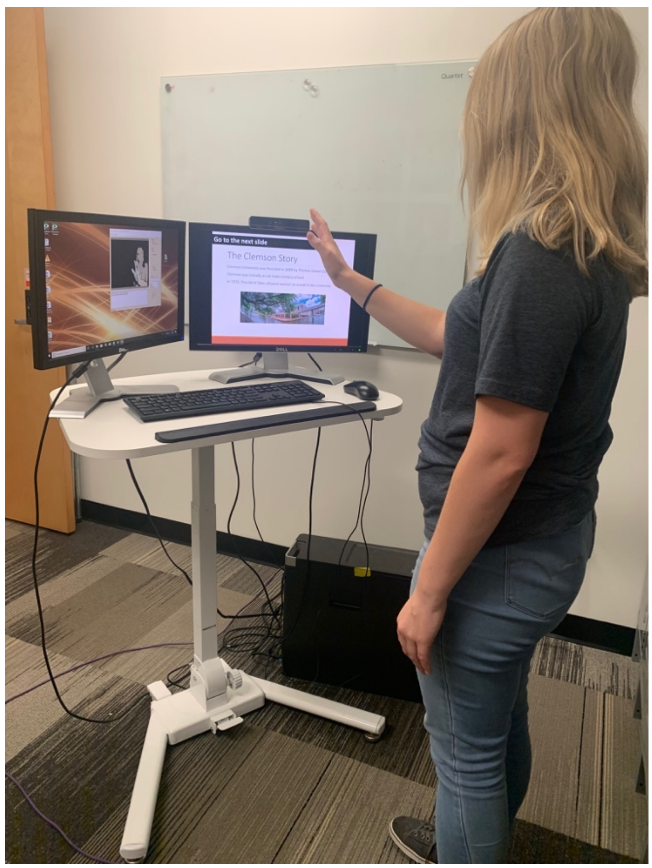


Figure 18. Experimental Setup for exposure and workload gestural study

Procedure

Upon arrival at the first session, the informed consent process was completed and the participant filled out a demographics survey and the Complacency Potential Rating Scales (Singh et al., 1993). Then, the researcher and the participant discussed how the camera worked and gestural interactions. The participant then completed the experimental task. The experiment incorporated a "Wizard of Oz" technique as done in previous studies (Jurewicz & Neyens, 2017; Jurewicz et al., 2018) and as described in Chapter 3. The function display (the effect) was always presented first and then the participant would choose a gesture (the cause) that they believed initiated the function. Participants performed gestures of their choosing and whichever gesture was their "first guess" to complete the function.

High workload participants were additionally instructed on performing an improvised speech on that session's topic (i.e., session 1 – speech on themselves, session 2 – speech on daily schedule, session 3 – speech on weekend plans). If these participants stopped their speech at any moment, the researcher would prompt them with questions in relation to the session's topic.

All participants completed an informal interview on their gesture choice at the end of each session and completed the User Acceptance Survey (Davis, 1989). Participants were asked questions in an informal interview related to the process behind generating gestures for particular functions and what previous experiences may have influenced their gesture choice. The researcher took notes during the interview and used comments as supportive qualitative findings to the intuitive gesture mappings.

Intuitive Gestures Measure

The intuitive gesture-function mappings were analyzed separately for the high workload and low workload conditions to identifying differences in mappings due to workload. The intuitive gesture-function mappings were additionally analyzed separately for each session to identify differences due to exposure. Videos of the participants' hands and fingers were recorded and analyzed to determine which gestures were performed for each function. The gestures were classified according to a gesture dictionary used in previous studies (Jurewicz & Neyens, 2017; Jurewicz et al., 2018) as well as Chapter 3. Two researchers independently classified gestures, and then consensus building was practiced to come to a consensus on gestures for all participants. According to Nielsen et al.'s (2004) approach, the intuitive gesture for a function is the gesture that is most frequently chosen across a group. The gesture responses for each function were compiled in a table, and the gesture response that was performed most frequently across the experimental group was chosen as the intuitive gesture-function mapping. This was done for each of the sessions and each workload condition.

Reaction Time Measure

The reaction time from presentation of the function display to completion of a gesture was recorded for every gesture-function pair by a Visual Basic program embedded in PowerPoint. The reaction time data was combined into one analysis for all sessions and both workload conditions. A mixed linear regression model with participant ID and block number as random effects was used to identify differences due to exposure or workload. A mixed linear regression model was used in order to account for both fixed and random effects. The fixed effects in the model were function, workload level, session, and responses from the demographic survey (e.g., , gender, race, handedness, education (highest degree obtained), video game use, and experience with virtual reality gaming.). The responses of variable “Video Game Use” were collapsed into two categories: “Yes” to playing video games and “No” to not playing video games. The “Yes” category included all positive responses to the video game use question from the demographic survey, and the “No” category included the negative response of “Do not play” video games.

The same formula for the mixed linear model from Chapter 3 was used in this analysis. Similarly to Chapter 3, ANOVAs were performed to determine if the random effects model was necessary for both the “block” random effect and “participant ID” random effect. Diagnostic tests were performed to ensure the assumptions for the linear model are met: linearity, homescedacity, normality, independence, and no multicollinearity issue

Tukey contrasts were calculated to make the pairwise comparisons for variables which had more than two level (i.e., differences between functions and sessions). R version 3.5.2 was used for all data analysis; the *lmer* function of the lme4 package (Bates et al., 2014) was used to build the mixed linear regression model, the ggplot2 package (Wickham, 2009) was used to plot the data, the *skewness* function of the e1071 package (Meyer et al., 2015), the *lmer* function of the lme4 package (Bates et al., 2014), the *glht* function of the multcomp package (Hothorn et al., 2008), and the *cooks.distance* function of the influence.ME package (Nieuwenhuis et al., 2012).

Results

The characteristics of the study participants are described in Table 4. The mean reaction time for all sessions and both levels of workload was 3.32 seconds (*SD*=1.97 seconds). Each participant had the opportunity to make 81 gestures, thus there were 3,240 possible gestures in this study. There were some participants that did not perform a gesture for a function, thus 3,222 gestures were analyzed. Overall, there were 30 unique gestures performed across all participants.

Table 4. Characteristics of participants in exposure and workload study

Variable Name	**N (%)**
Age	M=23.2, SD=5.06
Gender	
Male	10 (25)
Female	30 (75)
Handedness	
Right	37 (92.5)
Left	2 (5.0)
Ambidextrous	1 (2.5)
Education, highest degree obtained	
High School/GED	22 (55.0)
Bachelors	11 (27.5)
Masters	7 (17.5)
Video Game Use	
Yes	18 (45)
No	22 (55)
Virtual Reality Gaming Experience	
Yes	22 (55)
No	18 (45)

Intuitive Gesture Mapped

Bar graphs for gestures chosen across all participants are shown in Appendix C. Appendix D shows side by side comparisons of low vs. high workload mappings respectively, and Appendix E, F, and G show the intuitive mappings for session 1, 2 and 3 mappings, respectively. Pictorial representations of the gestures are provided in Figure 19.

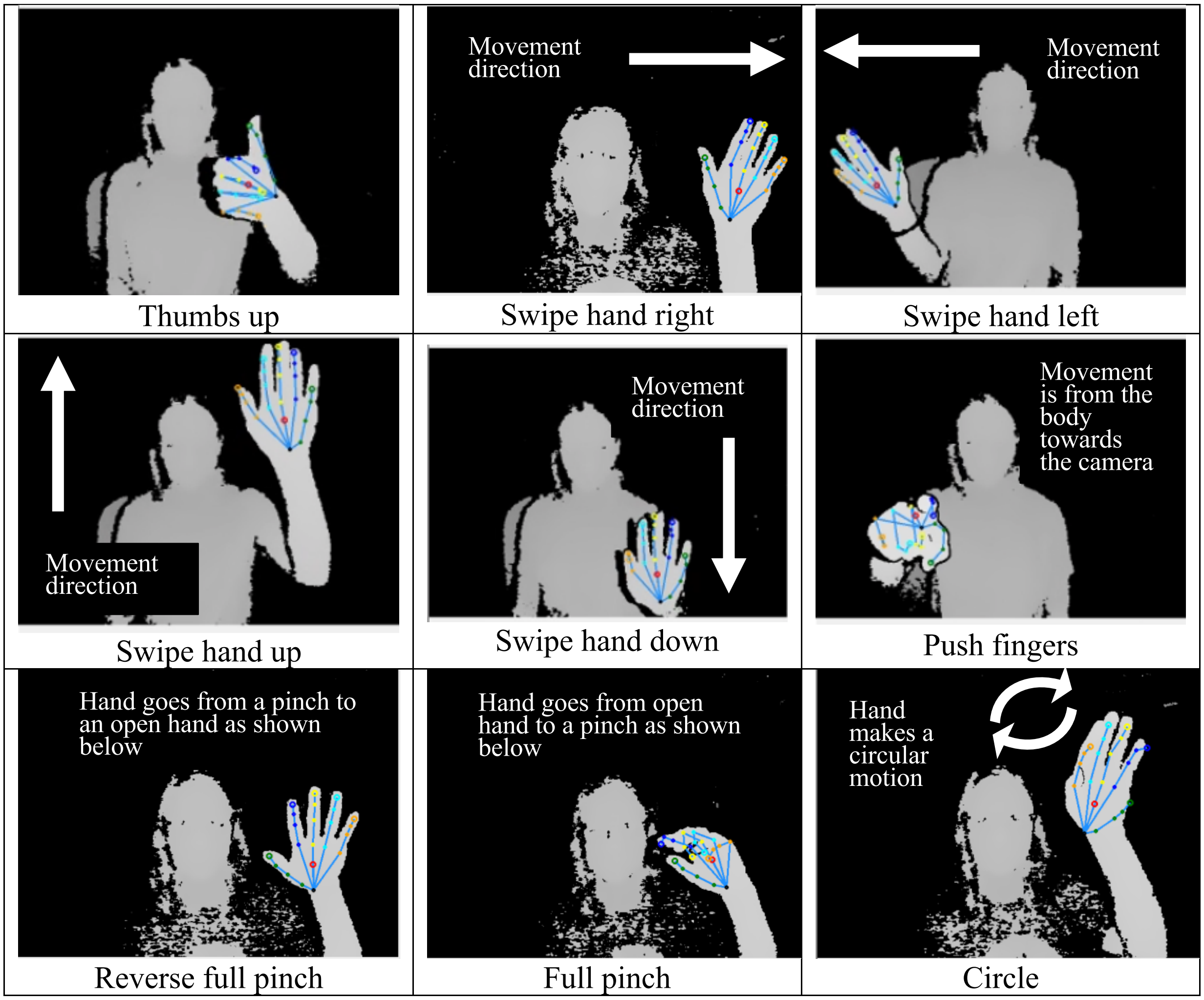


Figure 19. Pictorial representation of gestures from workload/exposure experiment

Table 5 shows that there are no differences to gesture-function mappings due to workload condition. “Swipe hand left” (*n*=82 for low, *n*=80 for high) was mapped to Function 1 (“Go to the next slide”), “Swipe hand left” (*n*=80 for low, *n*=69 for high) was mapped to Function 2 (“Go to the previous slide”), “Reverse full pinch” (*n*=173 for low, n=147 for high) was mapped to Function 3 (“Zoom in on the image”), “Full pinch” (*n*=100 for low, *n*=121 for high) was mapped to Function 4 (“Zoom out on the image”), “Swipe hand up” (*n*=64 for low, *n*=79 for high) was mapped to Function 5 (“Increase the volume”), “Swipe hand down” (*n*=81 for low, *n*=62 for high) was mapped to Function 6 (“Decrease the volume”), “Push fingers” (*n*=148 for low, *n*=170 for high) was mapped to Function 8 (“Select ‘College of Business’ from the menu”), and “Circle” (*n*=78 for low, *n*=77 for high) was mapped to Function 9 (“Undo the action”).

It was unclear for Function 7 (“Acknowledge the message”) what the intuitive mapping was because the top two gestures were performed almost equally as frequently. “Thumbs up” (*n*=59) and “Push fingers” (*n*=57) were performed most frequently for the low workload condition, and “Push fingers” (*n*=66) and “Thumbs up” (*n*=62) were performed most frequently for the high workload condition. Thus, Function 7 is shown to have two intuitive mappings for both the high and low workload conditions in Table 5.

Table 5. Intuitive gesture function mappings for workload level

Function [#]	Intuitive Gesture Mapped	
	Low Workload	**High Workload**
Go to the next slide [1]	Swipe hand left	Swipe hand left
Go to the previous slide [2]	Swipe hand right	Swipe hand right
Zoom in on the image [3]	Reverse full pinch	Reverse full pinch
Zoom out on the image [4]	Full pinch	Full pinch
Increase the volume [5]	Swipe hand up	Swipe hand up
Decrease the Volume [6]	Swipe hand down	Swipe hand down
Acknowledge the message [7]	Thumbs up/Push Fingers	Push fingers/Thumbs up
Select "College of Bus." [8]	Push fingers	Push fingers
Undo the action [9]	Circle	Circle

Table 6 shows that there are no differences to gesture-function mappings due to session. “Swipe hand left” (*n*=56, *n*=55, and *n*=51 for session 1, 2, and 3) was mapped to Function 1 (“Go to the next slide”), “Swipe hand left” (*n*=53, *n*=46, and *n*=50 for session 1, 2, and 3) was mapped to Function 2 (“Go to the previous slide”), “Reverse full pinch” (*n*=103, *n*=107, and *n*=110 for session 1, 2, and 3) was mapped to Function 3 (“Zoom in on the image”), “Full pinch” (*n*=67, *n*=73, and *n*=81 for session 1, 2, and 3) was mapped to Function 4 (“Zoom out on the image”), “Swipe hand up” (*n*=47, *n*=48, and *n*=48 for session 1, 2, and 3) was mapped to Function 5 (“Increase the volume”), “Swipe hand down” (*n*=48, *n*=48, and *n*=47 for session 1, 2, and 3) was mapped to Function 6 (“Decrease the volume”), “Push fingers” (*n*=104, *n*=106, and *n*=108 for session 1, 2, and 3) was mapped to Function 8 (“Select ‘College of Business’ from the menu”), and “Circle” (*n*=44, *n*=58, and *n*=53 for session 1, 2, and 3) was mapped to Function 9 (“Undo the action”).

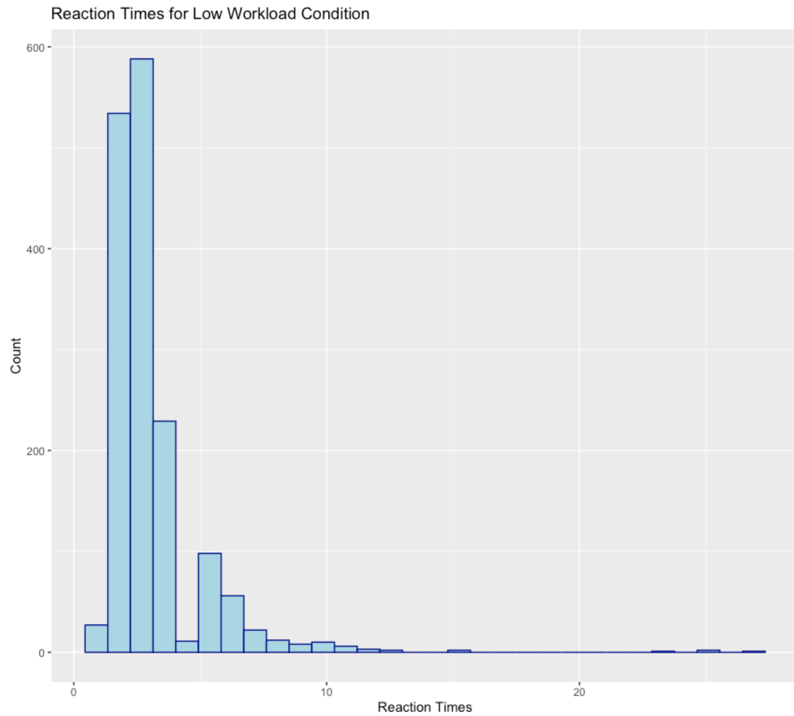
It was unclear for Function 7 (“Acknowledge the message”) what the intuitive mapping was because the top two gestures were performed almost equally as frequently. “Push fingers” (*n*=41) and “Thumbs up” (*n*=41) were performed most frequently during the first session, “Push fingers” (*n*=40) and “Thumbs up” (*n*=40) were performed most frequently during the second session, and “Push fingers” (*n*=42) and “Thumbs up” (*n*=40) were performed most frequently during the third session. Thus, Function 7 is shown to have two intuitive mappings for both the high and low workload conditions in Table 6.

Table 6. Intuitive gesture function mappings for each session

Function [#]	Intuitive Gesture Mapped		
	Session 1	**Session 2**	**Session 3**
Go to the next slide [1]	Swipe hand left	Swipe hand left	Swipe hand left
Go to the previous slide [2]	Swipe hand right	Swipe hand right	Swipe hand right
Zoom in on the image [3]	Reverse full pinch	Reverse full pinch	Reverse full pinch
Zoom out on the image [4]	Full pinch	Full pinch	Full pinch
Increase the volume [5]	Swipe hand up	Swipe hand up	Swipe hand up
Decrease the Volume [6]	Swipe hand down	Swipe hand down	Swipe hand down
Acknowledge the message [7]	Push fingers/Thumbs up	Push fingers/Thumbs up	Push fingers/Thumbs up
Select "College of Bus." [8]	Push fingers	Push fingers	Push fingers
Undo the action [9]	Circle	Circle	Circle

Reaction Times

The mean reaction time for the high and low workload conditions were 3.36 seconds (S*D*=1.96 seconds) and 3.29 seconds (*SD*=1.97 seconds), respectively. The mean reaction times for session 1, session 2, and session 3 were 4.11 seconds (*SD*=2.60 seconds), 3.12 seconds (*SD*=1.57 seconds), and 2.76 seconds (*SD*=1.20 seconds). Figures 20 and 21 show side by side histograms of reaction time data by workload level and session, respectively.



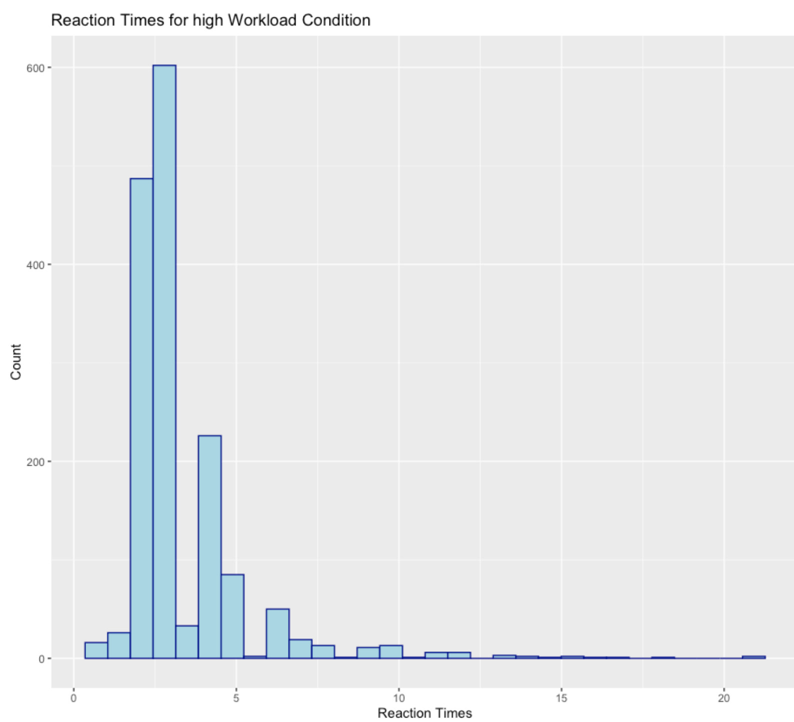


Figure 20. Histogram of reaction times for high and low workload conditions

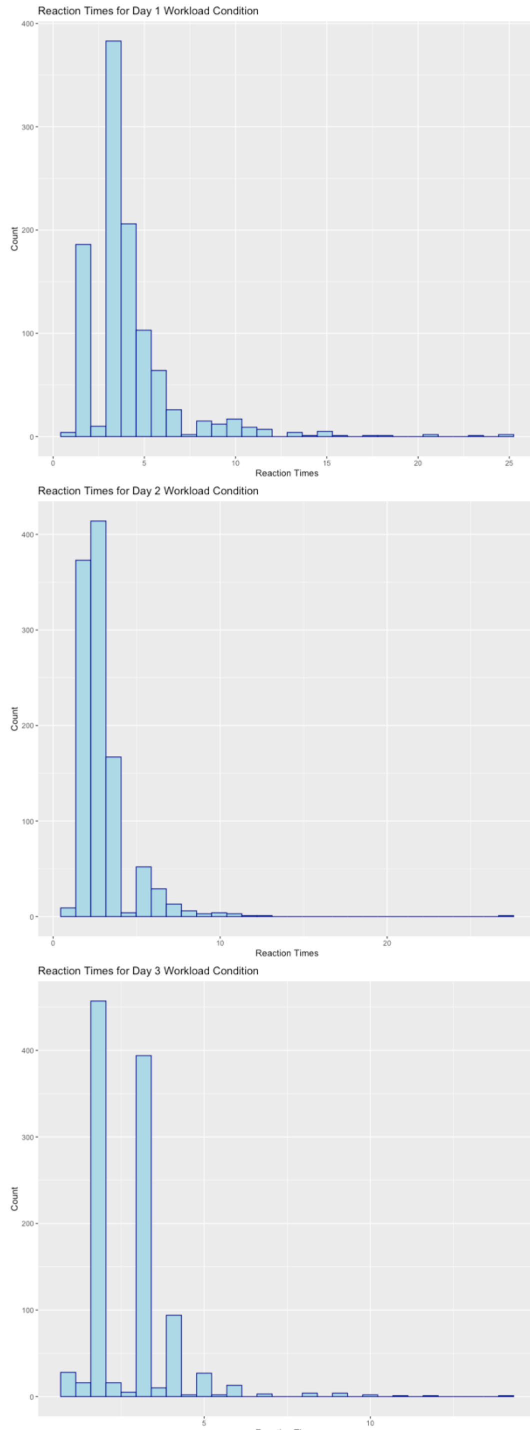


Figure 21. Histograms of reaction times for sessions 1-3

There were several assumptions that needed to be met before moving forward with the mixed model regression analysis, including normality, testing random effects vs. fixed effects model, linearity, independence, homoscedasticity, and no multicollinearity issues. The skew of response times exhibited a positive skew with a value of 4.15, so the data was transformed by taking the natural logarithm of response times. After the transformation, the skew was 0.86. The ANOVA of the model comparison showed that the linear regression model with “Participant ID” and “Block” as the random effects performed significantly better than the regression model with only the fixed intercept ($P<0.0001$), the regression model with only “”Participant ID” as a random effect ($P<0.0001$), and the regression model with only “Block” as a random effect ($P<0.0001$); therefore, the mixed model with two random effects, “Participant ID” and “Block” was used for the analysis.

None of the demographic variables were found to be significantly associated with reaction time, thus the final model only included function, workload, and session as predictor variables. The assumption of independence of the data was met in the final model because the within-subjects variability is removed as each participant is treated as a random effect and blocks are treated as a random effect. The VIF values of this model were calculated and all VIF values were less than 5 indicating that there were no severe multicollinearity issues. There were no influential points in the data set as calculated by Cook’s distance. The assumptions of linearity and homoscedasticity were met as confirmed by the residuals plot. The normal Q-Q plot confirmed normality of the response time data.

Table 7 shows that there is not a significant difference in reaction time with respect to workload. For function and session, Tukey's contrasts were calculated to perform pairwise comparisons among all levels of function and session. Table 8 shows the results for function contrasts and Table 9 shows the results for session contrasts. None of the functions were significantly different. However, all of the session days were significantly different in all pairwise comparisons ($p<0.0001$). The last session (session 3) was significantly faster than session 2 and session 1. Session 2 was significantly faster than session 1.

Table 7. Regression output of mixed effects linear model with exposure and workload

	β	*SE*	*t*	$P >\| t \|$
(Intercept)	1.272	0.077	16.438	<0.001*
Go to the previous slide [2]	0.016	0.024	0.661	0.509
Zoom in on the image [3]	0.041	0.024	1.676	0.094
Zoom out on the image [4]	0.014	0.024	0.570	0.569
Increase the volume [5]	0.074	0.024	3.035	0.002*
Decrease the Volume [6]	0.041	0.024	1.676	0.094
Acknowledge the message [7]	0.030	0.024	1.214	0.225
Select "College of Bus." [8]	0.008	0.024	0.334	0.738
Undo the action [9]	0.022	0.024	0.912	0.362
Session 2	-0.237	0.014	-16.815	<0.001*
Session 3	-0.346	0.014	-24.579	<0.001*
Low Workload	-0.016	0.011	-1.378	0.168

Table 8. Tukey contrasts for all function levels

Function Comparison	β	*SE*	*z*	$P > \lvert z \rvert$
2-1	0.0161	0.0243	0.6610	0.9992
3-1	0.0408	0.0243	1.6760	0.7613
4-1	0.0139	0.0243	0.5700	0.9997
5-1	0.0738	0.0243	3.0350	0.0607
6-1	0.0408	0.0243	1.6760	0.7613
7-1	0.0295	0.0243	1.2140	0.9536
8-1	0.0081	0.0243	0.3340	1.0000
9-1	0.0222	0.0243	0.9120	0.9924
3-2	0.0247	0.0243	1.0160	0.9845
4-2	-0.0022	0.0243	-0.0910	1.0000
5-2	0.0578	0.0243	2.3740	0.2982
6-2	0.0247	0.0243	1.0150	0.9845
7-2	0.0135	0.0243	0.5540	0.9998
8-2	-0.0079	0.0243	-0.3260	1.0000
9-2	0.0061	0.0243	0.2510	1.0000
4-3	-0.0270	0.0243	-1.1080	0.9733
5-3	0.0330	0.0244	1.3560	0.9140
6-3	0.0000	0.0243	-0.0010	1.0000

7-3	-0.0113	0.0244	-0.4630	0.9999
8-3	-0.0327	0.0244	-1.3410	0.9188
9-3	-0.0186	0.0243	-0.7650	0.9977
5-4	0.0600	0.0243	2.4680	0.2479
6-4	0.0269	0.0243	1.1070	0.9734
7-4	0.0157	0.0243	0.6460	0.9993
8-4	-0.0057	0.0243	-0.2350	1.0000
9-4	0.0083	0.0243	0.3430	1.0000
6-5	-0.0331	0.0243	-1.3590	0.9132
7-5	-0.0443	0.0243	-1.8210	0.6687
8-5	-0.0657	0.0243	-2.7010	0.1470
9-5	-0.0517	0.0243	-2.1230	0.4573
7-6	-0.0112	0.0243	-0.4620	0.9999
8-6	-0.0326	0.0243	-1.3410	0.9191
9-6	-0.0186	0.0243	-0.7640	0.9978
8-7	-0.0214	0.0243	-0.8800	0.9940
9-7	-0.0074	0.0243	-0.3020	1.0000
9-8	0.0141	0.0243	0.5780	0.9997

Table 9. Tukey contrasts for all session levels

Session Comparison	β	*SE*	*z*	$P >\mid z \mid$
2-1	-0.236	0.0141	-16.82	<0.001*
3-1	-0.345	0.0141	-24.58	<0.001*
3-2	-0.109	0.0140	-7.8	<0.001*

Discussion

The objective of this study was to examine the effect of workload and exposure on intuitive gesture-function mappings. There were 30 unique gestures generated in this study for Clemson-affiliated individuals controlling a PowerPoint presentation about Clemson with 3D gestures. Both high and low workload participants generated the same gesture-function mappings, and there was no significant difference in reaction times to generate an intuitive mapping between workload levels. All sessions generated generally the same gesture-function mappings; however, there were significant differences in reaction times between all pairwise comparisons of sessions with session 3 being the fastest and session 1 being the slowest to generate intuitive mappings.

Most of the functions that contrasted generated opposing gestures. For example, “Go to the next slide” and “Go to the previous slide” were mapped to “Swipe hand left” and “Swipe hand right,” respectively. The top four gestures for these two functions were “Swipe hand right,” “Swipe hand left,” “Swipe fingers right,” and “Swipe fingers left” showing that regardless of direction and hand posture (e.g., just fingers or whole hand), it was intuitive to participants to perform a lateral dynamic movement for going to the next and previous slides. Similarly, “Zoom in on the image” and “Zoom out on the image” were mapped to “Reverse full pinch” and “Full pinch;” and these gestures were the top two gestures performed for both functions indicating that regardless of direction, it was intuitive to the participant to perform a pinching motion. For “increase the volume” and “decrease the volume,” the top two gestures were “Swipe hand up”, ”Swipe fingers up” for increase and “Swipe hand down”, “Swipe fingers down” for decrease. This finding indicates that regardless of hand posture, it is intuitive to the participant to make a swiping motion either up or down for increase and decrease. It only happened once out of 3,222 times that a participant did an upward motion for decrease, and there were no downward motions generated for increase strongly suggesting the connection of increase to “up” and decrease to “down.” Many participants indicated in their qualitative responses that they found it easier to generate gestures for opposing functions.

The three remaining functions did not have contrasting actions in the function set, thus did not have opposing gesture mappings. “Acknowledge the message” did not necessarily have an intuitive mapping; however, a majority of participants either performed the dynamic gesture of “push fingers” or the static gesture of “thumbs up.” The two mappings may suggest that there may be multiple intuitive mappings for acknowledgement. “Select ‘College of Business’ from the menu” was intuitively mapped to “push fingers.” The second top gesture for this function, “push hand” only happened 11 times out of 3,222 opportunities, thus showing that the pushing motion was intuitive to the user. Function 9, “Undo the action,” was mapped to “Circle,” and participants verbally indicated that they were trying to imitate the “undo” icon on a computer by creating an in-air circle with their hand and fingers.

All participants were considered domain experts having used PowerPoint regularly and being familiar with Clemson University, so the result of exposure and workload not influencing intuitive gesture-function mappings shows that domain experts generate intuitive mappings at the first exposure to gestural control of a presentation even when in a high workload scenario. Workload also did not influence reaction times to generate an intuitive mapping potentially indicating that it was not more difficult in the higher workload scenario. However, there was an effect of exposure on reaction time. This potentially shows the learning effect of gestural interactions as the participants performed gestures faster and faster each session.

Although an effect of workload was not found in this study, it cannot be definitively concluded that there is not an effect of workload for gestural interactions. There were 20 participants in each workload condition, so sample size could be increased to generate more power in the study. However, 1,610 gestures were analyzed per workload condition, so the lack of a workload effect may actually suggest that the high workload condition was not difficult enough or there really is not an effect of workload in this context. Future work could investigate more difficult high workload conditions by having participants perform more difficult tasks while generating gestures.

CHAPTER 5: BAYESIAN STATISTICAL ANALYSIS OF EXPERTISE AND EXPOSURE IN 3D GESTURAL INPUT SYSTEMS

Introduction

The work in this chapter addresses the third research question of this dissertation: Q3 – How accurately can gestures be statistically predicted? A Bayesian analysis was performed for the expertise data from Chapter 3 as well as the exposure data from Chapter 4. The analyses are performed separately due to the differences in context. In Bayesian methods, Bayes theorem is utilized to obtain a posterior distribution by identifying the prior distribution, which represents prior beliefs, and the likelihood function that is based on the data structure. The posterior distribution is an update of the prior beliefs after seeing new data. If the complete posterior distribution is obtainable, then it provides all the information needed to make posterior inferences. However, it is often the case that the posterior distribution is unrecognizable and does not have a form that is easily sampled from. This chapter shows how a Bayesian analysis is performed on intuitive gestural input data (i.e., gesture choice) by first explaining the structure of the data and its impacts on the Bayesian analysis, showing results from two multinomial Bayesian logistic regression models, and discusses insights into the advantages of using Bayesian statistics for understanding human behavior. For simplicity and ease of understanding, the expertise Bayesian model will be known as Model 1 and exposure Bayesian model will be known as Model 2.

Data Structure and Approach

The exposure and expertise data are used in the Bayesian model to predict intuitive gesture choice based on the context and either expertise or exposure. Thus, the observed data, or the output of each model, is gesture choice as described by $y_1,\ldots,y_n$. Gesture choice is categorical with many categories, so the observed data is distributed as multinomial:

$$y_i, \ldots, y_n \sim Multinomial\ (n, p)$$

Where p is a vector representing the probabilities of each gesture category and n is the sample size. Suppose k represents the total number of gesture categories, so the p vector is represented as $p = (p_1, \ldots, p_k)$. The joint probability mass function (pmf) of the observed multinomial data, $y_i, \ldots, y_n$, with probabilities, $p = (p_1, \ldots, p_k)$, and a sample size of n is given by:

$$P(y|p) = \frac{n!}{\prod_{i=1}^{k} y_i} \prod_{i=1}^{k} {p_i}^{y_i} = \frac{n!}{y_1!\, y_2! \ldots y_k!} {p_1}^{y_1} {p_2}^{y_2} \ldots, {p_k}^{y_k}$$

Where $\sum_{i=1}^{k} y_i = n$ and $\sum_{i=1}^{k} p_i = 1$

The purpose of Bayesian methods is to obtain the posterior distribution. In both Model 1 and Model 2, form of the data (i.e., the likelihood function) for gesture choice is known, so the only thing missing is the prior distribution. It is of interest to identify a Bayesian posterior which is friendly to work with and sample from. Looking at the general Bayesian posterior, $P(\theta|y)$, one way to obtain a friendly posterior distribution is by using conjugate priors. The concept of conjugacy ensures that given the form of the data, the prior and posterior distributions come from the same distribution family. If conjugate priors are used, then all of the information about the posterior is known and integrals can be directly computed to obtain posterior estimates.

To identify the conjugate prior for the multinomial data in this dissertation, the form of the data (i.e., the likelihood) is used. The likelihood function refers to the pmf, therefore:

$$L(y|p) = p(y|p) = \frac{n!}{\prod_{i=1}^{k} y_i} \prod_{i=1}^{k} {p_i}^{y_i}$$

Since n is fixed, we really only care about p therefore we can simplify to:

$$L(y|p) \propto \prod_{i=1}^{k} {p_i}^{y_i}$$

We recognize this to follow the form of a kernel of the Dirichlet distribution:

$$P(p|\alpha) \propto \prod_{i=1}^{k} {p_i}^{a_i - 1}$$

Aside: the probability density function of the Dirichlet distribution is:

$$P(y|p) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} {p_i}^{a_i - 1}$$

Therefore, the conjugate prior of a multinomial distribution is a Dirichlet distribution. That is,

$$p \sim Dirichlet(a_1, \ldots, a_k)$$

Since we are using a conjugate prior, the posterior distribution will also be Dirichlet:

$$P(p|y) \propto P(y|p)P(p) = \left(\frac{n!}{\prod_{i=1}^{k} y_i} \prod_{i=1}^{k} {p_i}^{y_i}\right)\left(\frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} {p_i}^{a_i - 1}\right) \propto \prod_{i=1}^{k} {p_i}^{a_i + y_i - 1}$$

Therefore, the conjugate posterior of a multinomial distribution is Dirichlet:

$$p|y \sim Dirichlet(a_1 + y_1, \ldots, a_k + y_k)$$

A friendly posterior distribution can be obtained for multinomial data, but given the complexity behind direct integration of multinomial data, it is generally not preferred to perform integration. Instead, simulated draws can be taken from the posterior distribution, most popularly done by using Monte Carlo Markov Chains (MCMCs). In MCMC, a sampling strategy is setup that generates a markov chain in which the stationary distribution equals the posterior distribution of interest. However, deriving a posterior distribution that is easily sampled from is not always straightforward especially given multivariate problems with multiple covariates. One way that posterior estimates can still be obtained in complex problems is through the Metropolis-Hastings Algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953). Metropolis-Hastings is a general way of constructing a markov chain in which individual draws are proposed and the draws converge to the target distribution by using an acceptance/rejection rule. The Metropolis-Hastings algorithm performs the following:

Let: $p(\theta|y)$ be the target distribution

$\theta^{(t)}$ be the current draw from $p(\theta|y)$

$g\left(\theta\middle|\theta^{(t)}\right)$ be the proposal distribution

Metropolis-Hastings Steps:

1. Propose $\theta^{*} \sim g\left(\theta\middle|\theta^{(t)}\right)$
2. Calculate Metropolis-Hastings ratio

$$\alpha = min\left\{1, \frac{p(\theta^{*}|y)g\left(\theta^{(t)}\middle|\theta^{*}\right)}{p\left(\theta^{(t)}\middle|y\right)g\left(\theta^{*}\middle|\theta^{(t)}\right)}\right\}$$

3. Accept $\theta^{(t+1)} = \theta^{*}$ with probability α. Otherwise, set $\theta^{(t+1)} = \theta^{(t)}$

The ratio in Step 2 is the Metropolis-Hastings ratio and essentially calculates whether a proposed draw, θ^{*}, is more likely (i.e., has a higher density) than the current draw, $\theta^{(t)}$. The proposed draw is accepted as the new draw with a probability of α, and if the proposed draw is not accepted, then the new draw remains the current draw. This process is repeated until enough iterations of the algorithm have completed such that the chain has converged.

Metropolis-Hastings is highly versatile and is used widely in Bayesian statistics so that accurate posterior estimates can still be obtained even when a posterior distribution is unrecognizable.

Bayesian Multinomial Logistic Regression

Overall, the goal of this dissertation is to accurately predict intuitive gesture choice, thus, it is a classification problem. The data is multinomial, and simple Bayesian analyses could be performed on the expertise and exposure data sets with the Multinomial-Dirichlet conjugate priors. If conjugate priors are used to perform Bayesian classification for gesture choice, no other information could be included in the analysis such as the contextual task (i.e., the function), expertise, or exposure. The Bayesian classification results with conjugate priors would be relatively uninformative as it would only provide overall information on the likelihood of each gesture choice. Instead, it is of interest to understand how individual or contextual factors influence the probability of particular gesture choices.

This dissertation seeks to model the probability that an observation, y_i, takes a certain gesture category where:

K is the set of all possible gesture choices

$P(y_i = k), \forall k \in K$

$\boldsymbol{x}$ represents a covariate vector

The covariates in this case are contextual task, expertise level, and exposure time. Thus, this data can be modelled as a general regression problem where $\boldsymbol{\beta}$ represents the unknown regression coefficients:

$$\boldsymbol{y_i} = \boldsymbol{x\beta} + \boldsymbol{\varepsilon}$$

Given that the multinomial data is to be modelled as a regression problem, the general regression model translates to the Bayesian format as a Bayesian Multinomial Logistic Regression:

$$P(y_i = k|\boldsymbol{\beta}) = \frac{\exp\,(\boldsymbol{x}_{ki}\boldsymbol{\beta})}{1 + \sum_{j=1}^{m} \exp\,(\boldsymbol{x}_{ki}\boldsymbol{\beta})}$$

Where $p(\boldsymbol{\beta}|\boldsymbol{y})$ is the posterior distribution. Thus, the aim is to obtain posterior estimates from $p(\boldsymbol{\beta}|\boldsymbol{y})$ given some covariate information. The regression coefficients, $\boldsymbol{\beta}$ **,** depend on the contextual task and the user group or exposure time, so there is not friendly way to sample from the posterior directly. Therefore, other sampling variations must be used to obtain posterior estimates, such as variations of the Metropolis-Hasting algorithm.

Model 1: Expertise Data

Model Fitting and Diagnostics

The expertise data from Chapter 3 is fit to a Bayesian Multinomial Logistic Regression model using the Metropolis-Hastings algorithm. Gesture choice is modeled based on the contextual task and expertise (i.e., function and user type). A function was developed in R 3.6.1 adapted from the MCMCpack package (Martin, Quinn, & Park, 2011). A Random Walk Metropolis-Hastings algorithm was used, and the algorithm is exactly the same as the original but the proposal distribution is assumed to be symmetric (Robert, Elvira, Tawn, & Wu, 2018). The alternative is to perform an independent chain Metropolis-Hastings, but this approach is rarely used as it requires some amount of prior knowledge to build a proposal distribution that is relevant to the problem (Robert et al., 2018). The algorithm went through 1,000,000 iterations with a burn-in of 10,000 and a thinning parameter of 10. The thinning parameter reduces autocorrelations in the data, and in this case, every 10^{th} iteration of the chain is returned. The prior distribution for the regression coefficients, $\boldsymbol{\beta}$**,** was set to be distributed as a multivariate normal distribution (i.e., $\sim MVN(\boldsymbol{b}_0, \boldsymbol{B}_0^{-1})$). The hyperparameters of the MVN distribution were centered at zero with a very large variance so that the prior distribution is a non-informative, relatively flat prior that still has the integrity and characteristics of a MVN distribution. A tuning parameter was manipulated in the Metropolis-Hastings algorithm such that the acceptance rate was between 20% and 50% (Gelman et al., 2013).

Model 1 had an acceptance ratio of 26.502% with the tuning parameter set to 0.17. In all regression analysis, diagnostics need to performed to ensure that the model is a good fit to the data. In Bayesian regression analysis, trace plots and autocorrelation plots are two ways to identify a good model fit. A sample of the trace plots for the $\boldsymbol{\beta}$ are shown in Figure 22 and indicate that the chains converged and the chain has explored the full parameter space.

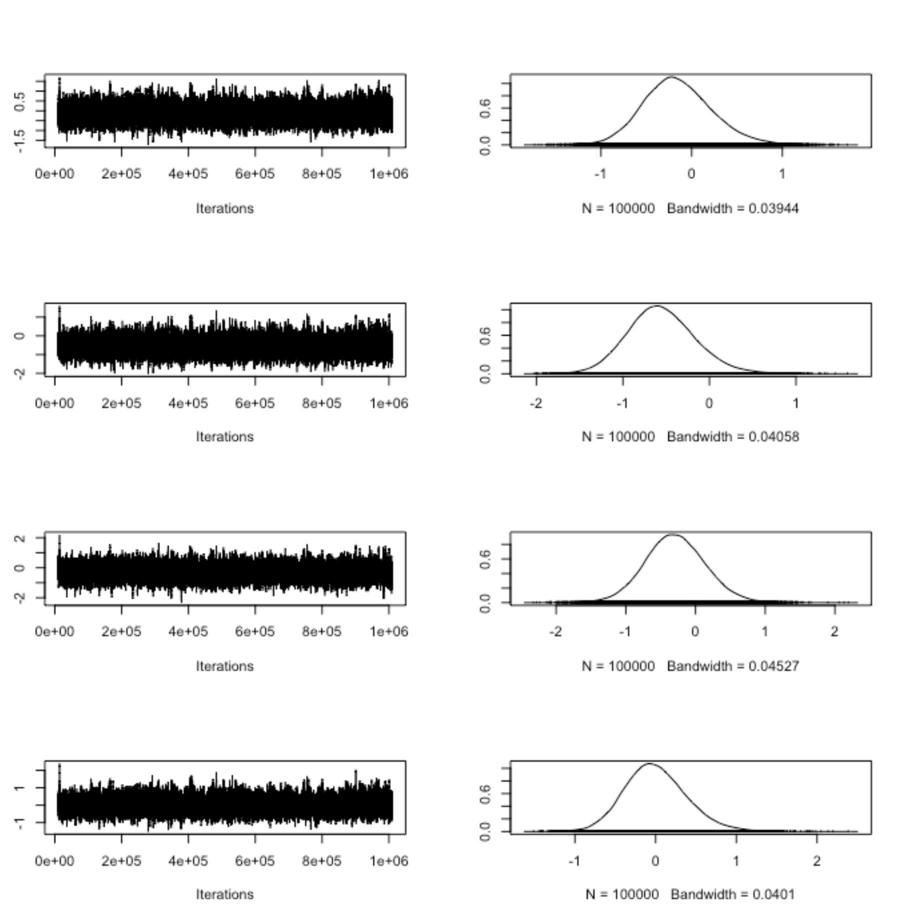


Figure 22. Sample of trace plots for regression coefficients for Model 1

A sample of the autocorrelations of $\boldsymbol{\beta}$ is shown in Figure 23. If there are large autocorrelations, then it would indicate that the chain is not mixing well and the chain has not explored the full space of the posterior distribution. Figure 23 shows that autocorrelation decreases rapidly as a function of lag until it is nearly zero.

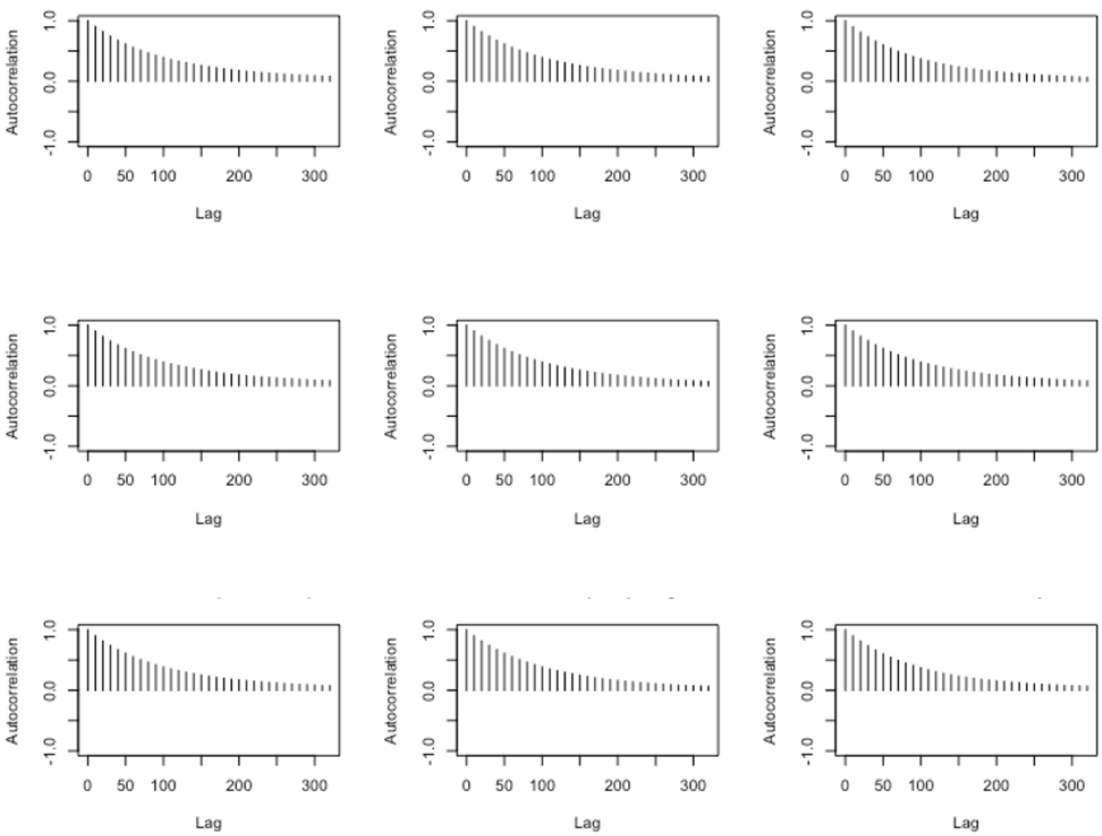


Figure 23. Sample of autocorrelations of regression coefficients for Model 1

Posterior Predictive Probabilities

The Bayesian Multinomial Model was fit for the expertise data and the diagnostics showed that the Metropolis-Hastings algorithm converged to the posterior distribution. The Bayesian model can be used to explain what someone has done, and with any regression problem, it can predict what the next person will do. Thus, predictions of future system use can be performed and a probability for the predictions are directly calculated. The results of the predictive probabilities for Model 1 is summarized in Table 10. The predicted gesture choice (i.e., the gesture with the highest probability) is provided for all combinations of functions and user type where the user type is either unspecified, anesthesia provider, or student.

Table 10. Posterior predictive probabilities of intuitive gesture choice based on expertise and the contextual task for Model 1

Function [#]	Predicted Gesture Choice		
	User Type Unspecified	**Novices**	**Experts**
Start the flow [1]	Swipe hand up (0.148)	Thumbs up (0.349)	Rotate right (0.348)
Stop the flow [2]	Swipe hand down (0.137)	Five up (0.164)	Rotate left (0.239)
Inc. the flow [3]	Swipe hand up (0.182)	Swipe hand up (0.182)	Rotate right (0.298)
Dec. the flow [4]	Swipe hand down (0.217)	Swipe hand down (0.208)	Rotate left (0.389)
Silence alarm [5]	Push hand (0.240)	Swipe hand left (0.223)	Push hand (0.255)
Ack. the message [6]	Thumbs up (0.251)	Thumbs up (0.226)	Thumbs up (0.302)
Heart rate normal? [7]	Thumbs up (0.248)	Thumbs up (0.219)	Thumbs up (0.326)
Pulse ox normal? [8]	Thumbs up (0.237)	Thumbs up (0.202)	Thumbs up (0.327)
Select heart rate [9]	Thumbs up (0.216)	Push fingers (0.176)	Push fingers (0.304)
Cancel the message [10]	Thumbs up (0.184)	Swipe hand left (0.137)	Swipe hand left (0.236)

Note: All values in the table represent the probability that y_{n+1} takes on a gesture, *k*, given the type of user

The predicted gesture choice is the same for all three user types for functions 6-10 and the predicted gesture choice is different for functions 1-5, similarly to the results from Chapter 3 which showed the same intuitive gesture choice differences via the frequency analysis. The Bayesian model not only provides the most probable gesture choice as in Table 10 but also gives the posterior predictive probabilities for all possible gesture choices. Figures 24-26 show an example of the plots of posterior predictive probabilities for "Function 1 – Start the Flow of Anesthesia Gas" for unspecified user type, anesthesia providers, and students, respectively.

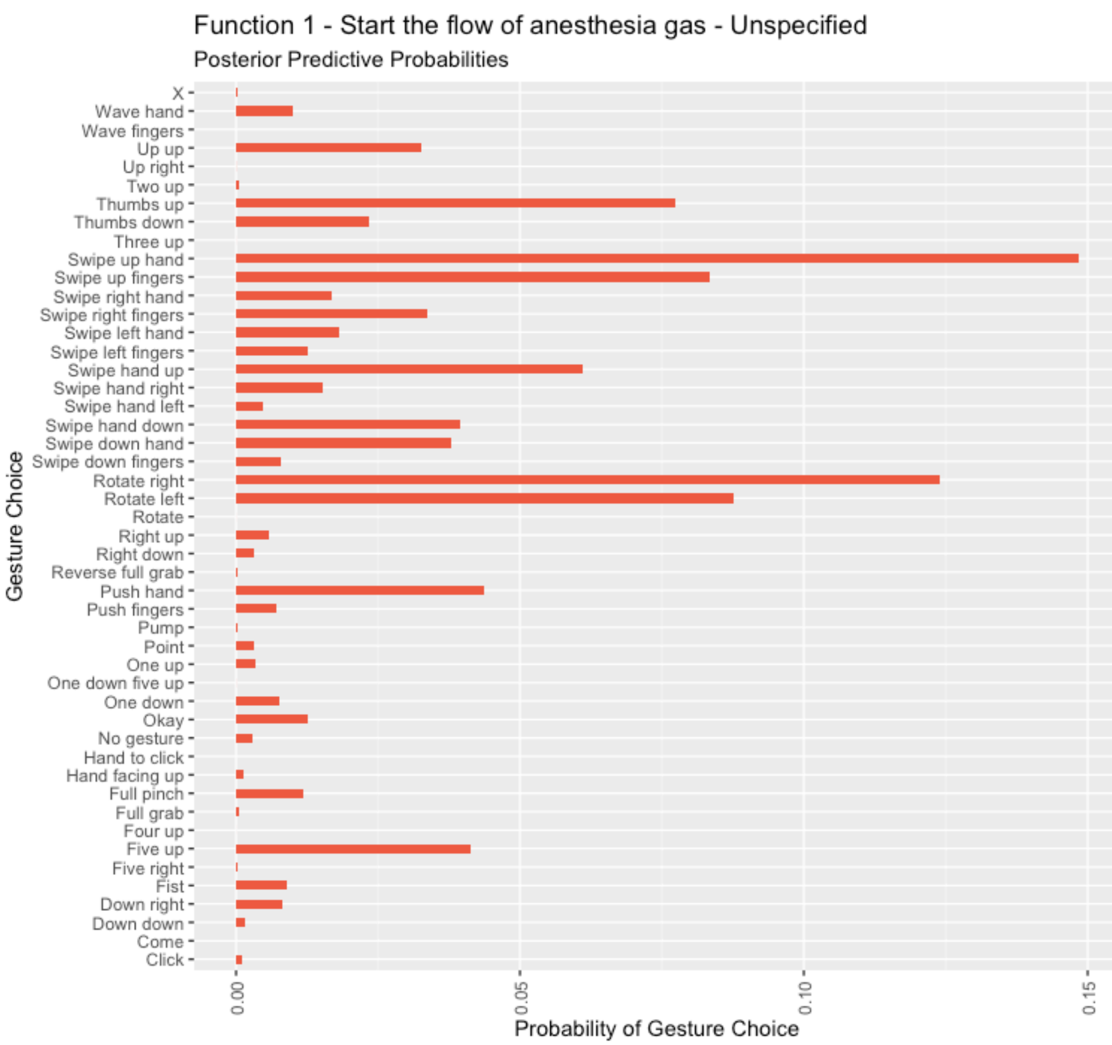


Figure 24. Predictive probabilities for all gesture choices for function 1 and user type unspecified.

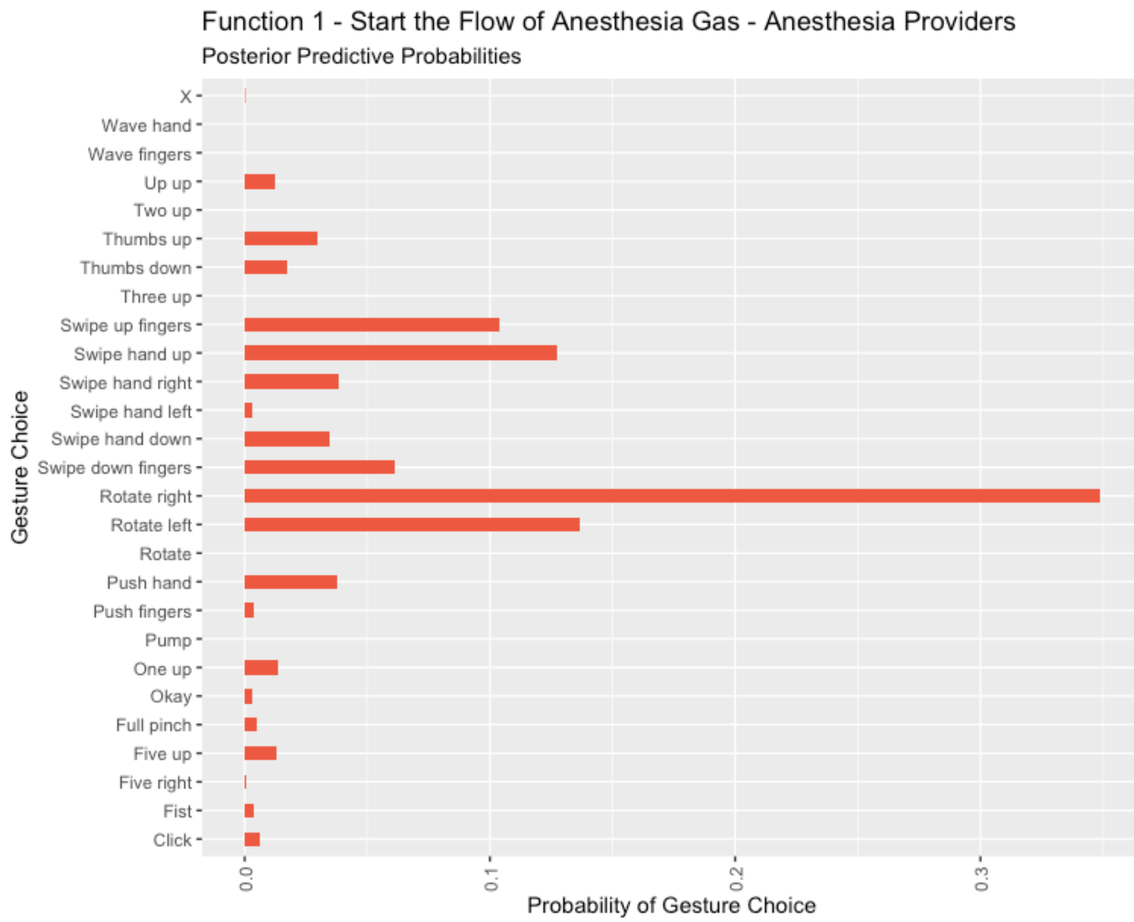


Figure 25. Predictive probabilities for all gesture choices for function 1 and user type specified as anesthesia provider

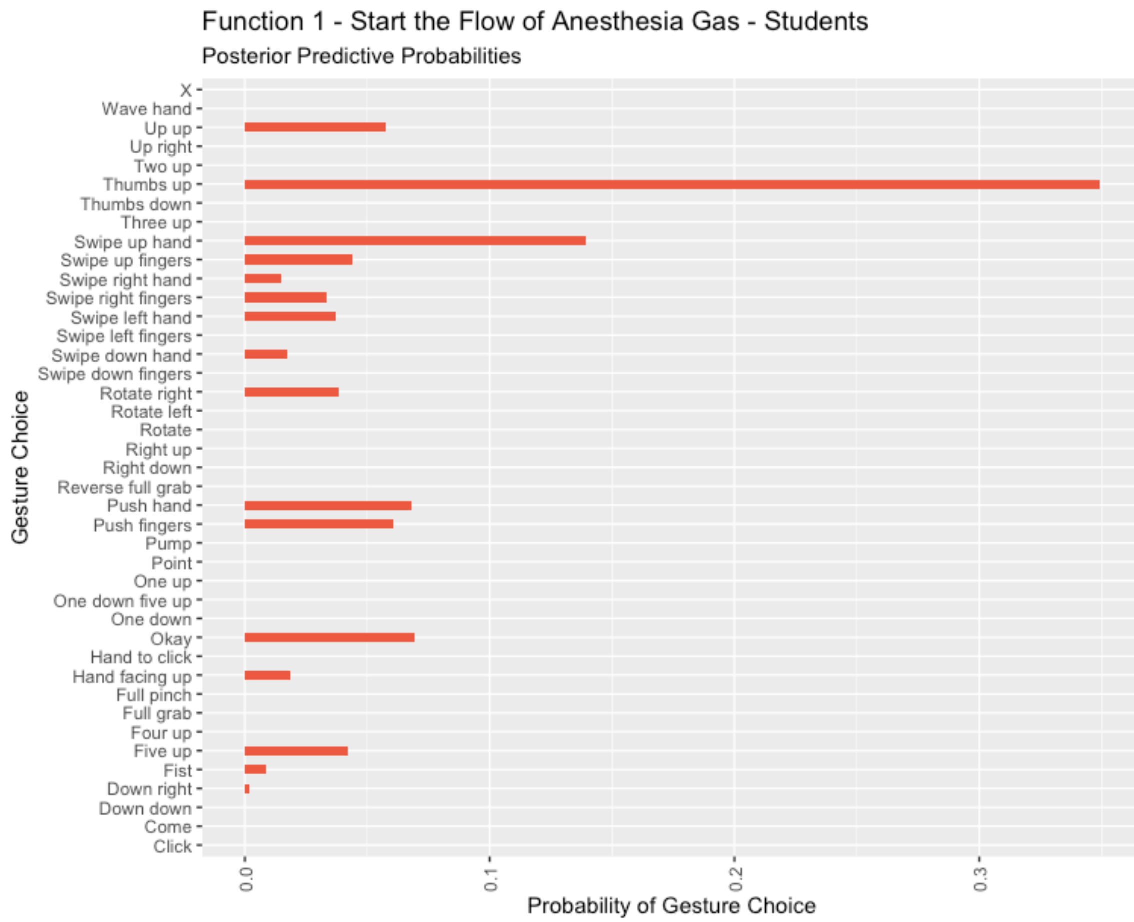


Figure 26. Predictive probabilities for all gesture choices for function 1 and user type specified as student

Figures 24 shows that when no information is known about the user, the best prediction that can be made for starting the flow of anesthesia gas is a “swipe hand up” gesture with a probability of 0.148. The next most probable predicted gesture is “rotate right” with a probability of 0.122. In Figure 25, it is shown that when it is known that the user is an anesthesia provider, the best prediction for starting the flow of anesthesia gas is a “rotate right” gesture with a probability of 0.348, an increase from when the user type is not specified. Additionally, the next most probable gesture is “rotate left” with less than a

0.15 probability. In Figure 26, it is shown that when the user is a student, the best prediction for starting the flow of anesthesia gas is a "thumbs up" gesture with a probability of 0.349, also an increase from when the user type is unspecified. The second most probable gesture for students is "swipe hand up" with a probability of about 0.15. The plots show similar patterns for the other functions which had differing predictive probabilities.

Some functions had the same predicted gestures despite specification of user type. Figures 27-29 show an example of the plots of posterior predictive probabilities for "Function 7 – Is heart rate normal?" for unspecified user type, anesthesia providers, and students, respectively.

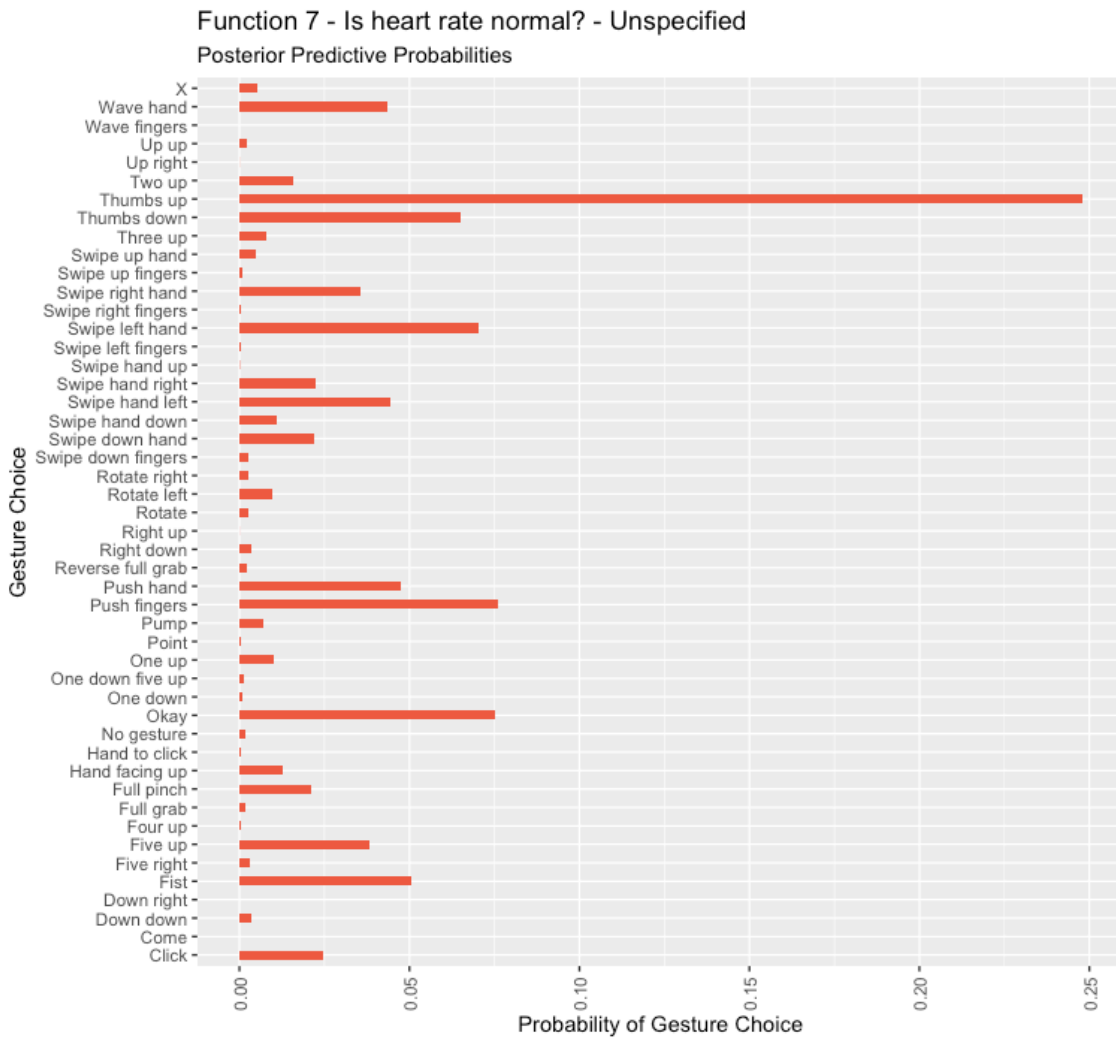


Figure 27. Predictive probabilities for all gesture choices for function 7 and user type unspecified.

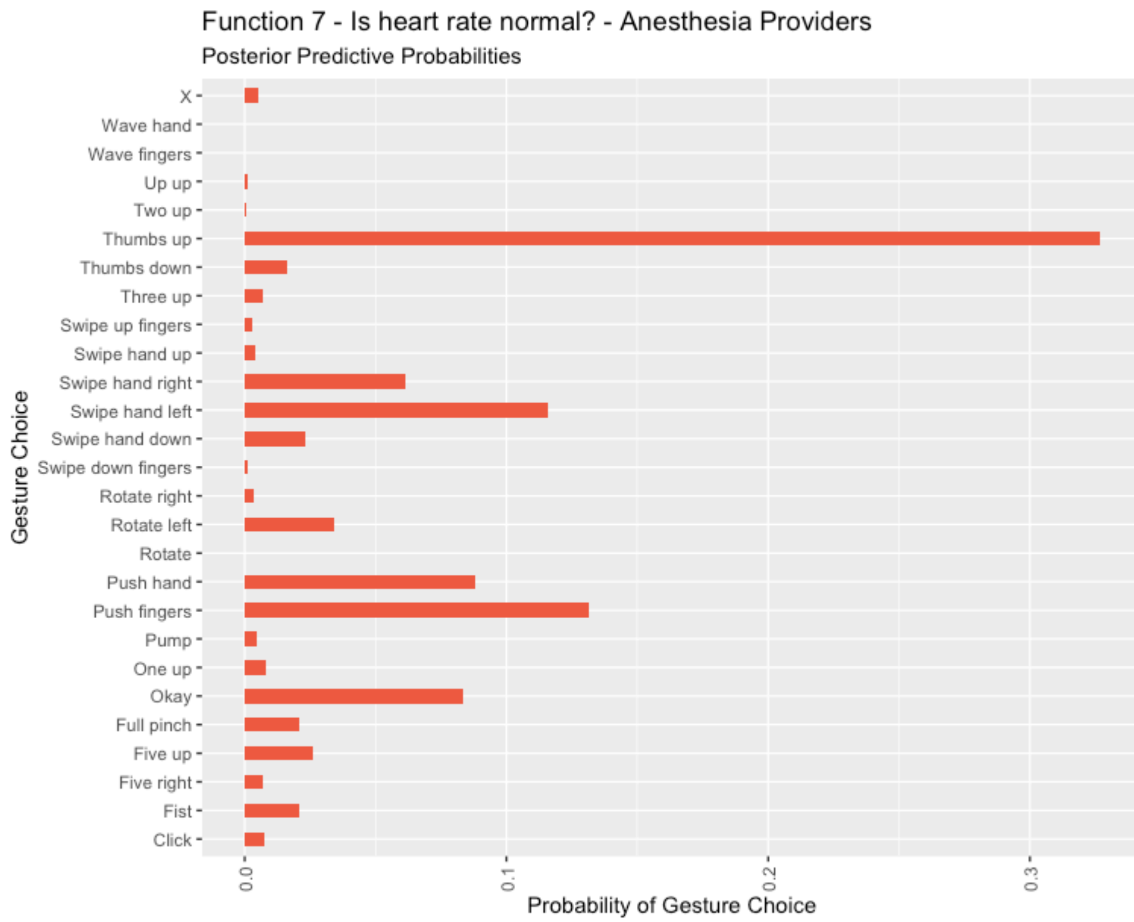


Figure 28. Predictive probabilities for all gesture choices for function 7 and user type specified as anesthesia provider

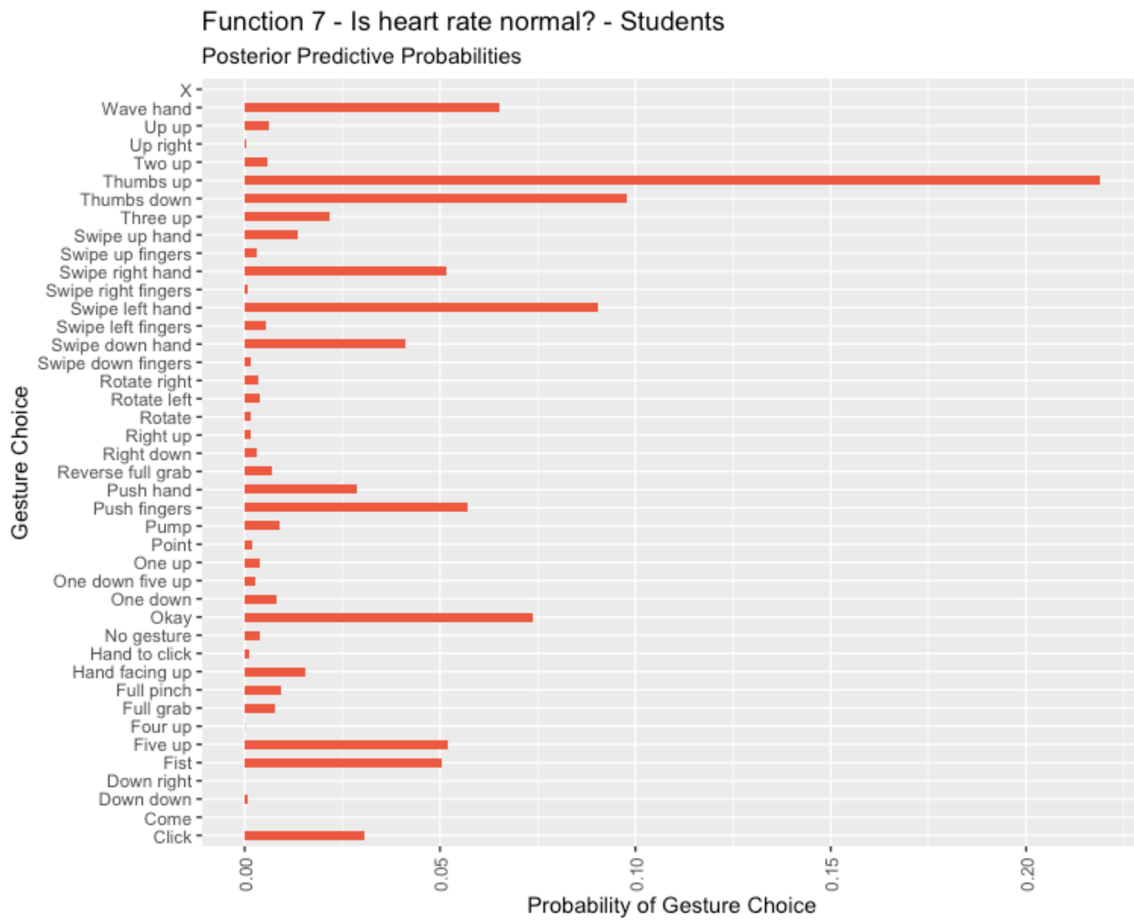


Figure 29. Predictive probabilities for all gesture choices for function 7 and user type specified as student

Figures 27-29 show that the most probable gesture to be performed by the next user is “thumbs up” with a probability of 0.248 for user type unspecified, 0.326 for anesthesia providers, and 0.219 for students. Students had the lowest predictive probability among all user types; however, the second most probable gesture is “thumbs down” with a probability of about 0.10. The correct assessment of “Function 7-Is heart rate normal?” was a “Yes” that the heart rate was normal, thus the predicted gesture may not be accurate due to lack of experience in the students understanding the heart rate parameter explaining why the student predictions were the lowest.

Model 2: Exposure Data

Model Fitting and Diagnostics

The exposure data from Chapter 4 is fit to a Bayesian Multinomial Logistic Regression model using the Metropolis-Hastings algorithm as previously done with Model 1. Gesture choice is modeled based on the contextual task and exposure (i.e., function and session number). The same R function from Model 1 was used to fit Model 2, and the same variations applied including the Random Walk Metropolis-Hastings algorithm, 1,000,000 iterations, a burn-in of 10,000, thinning parameter set to 10, and the tuning parameter set such that the acceptance rate was between 20% and 50% (Gelman et al., 2013). The prior distribution for the regression coefficients, $\boldsymbol{\beta}$**,** was also set to be distributed as a multivariate normal distribution (i.e., $\sim MVN(\boldsymbol{b}_0, \boldsymbol{B}_0^{-1})$) centered at zero with a very large variance so as to set the prior to be non-informative and relatively flat

Model 2 had an acceptance rate of 21.792% with the tuning parameter set to 0.31, and trace plots and autocorrelation plots were constructed to perform model diagnostics. A sample of the trace plots for the $\boldsymbol{\beta}$ are shown in Figure 30 and indicate that the chains converged and the chain explored the full parameter space.

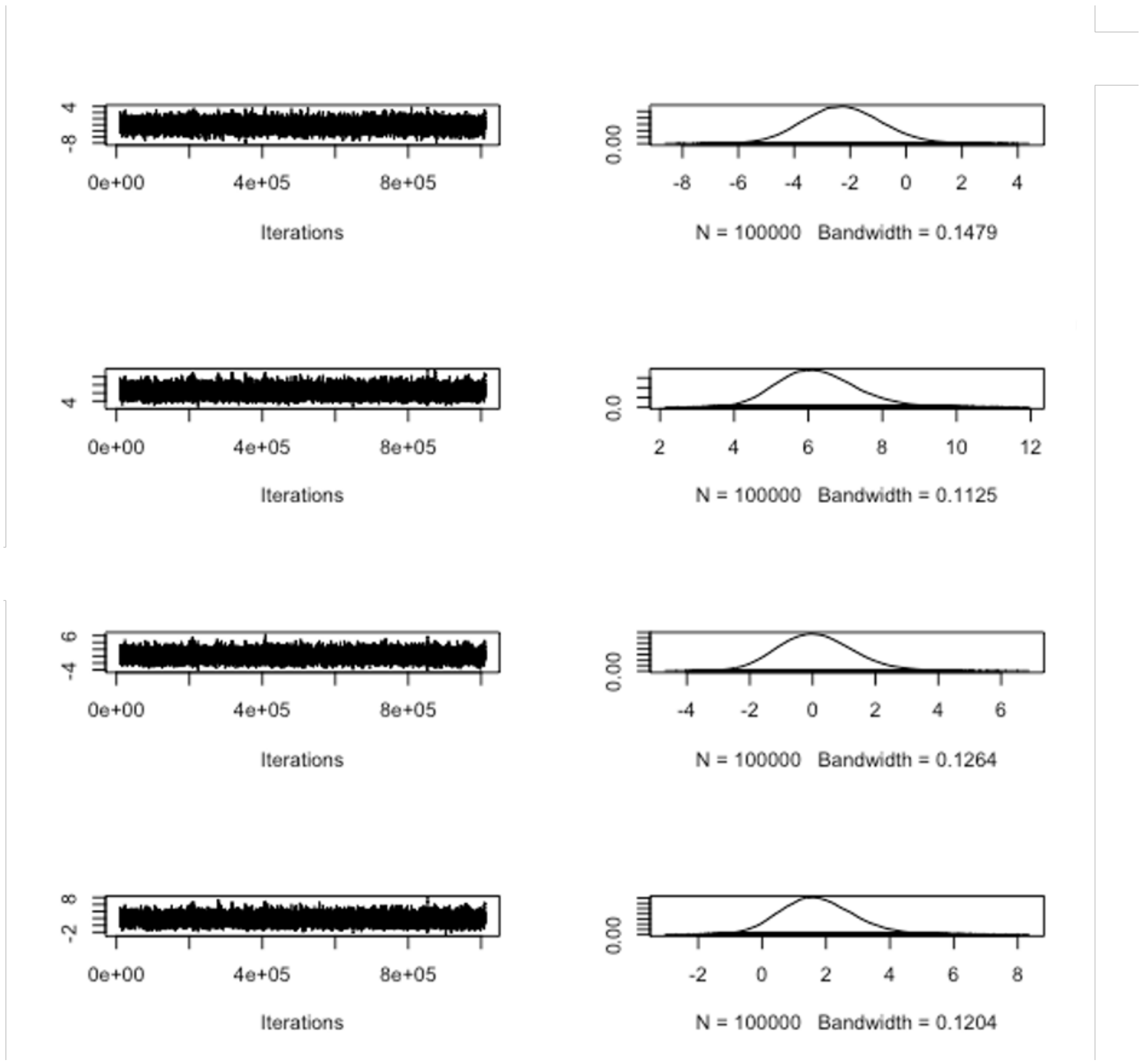


Figure 30. Sample of trace plots for regression coefficients for Model 2

A sample of the autocorrelations of $\boldsymbol{\beta}$ is shown in Figure 31 and shows that autocorrelation decreases rapidly as a function of lag until it is nearly zero.

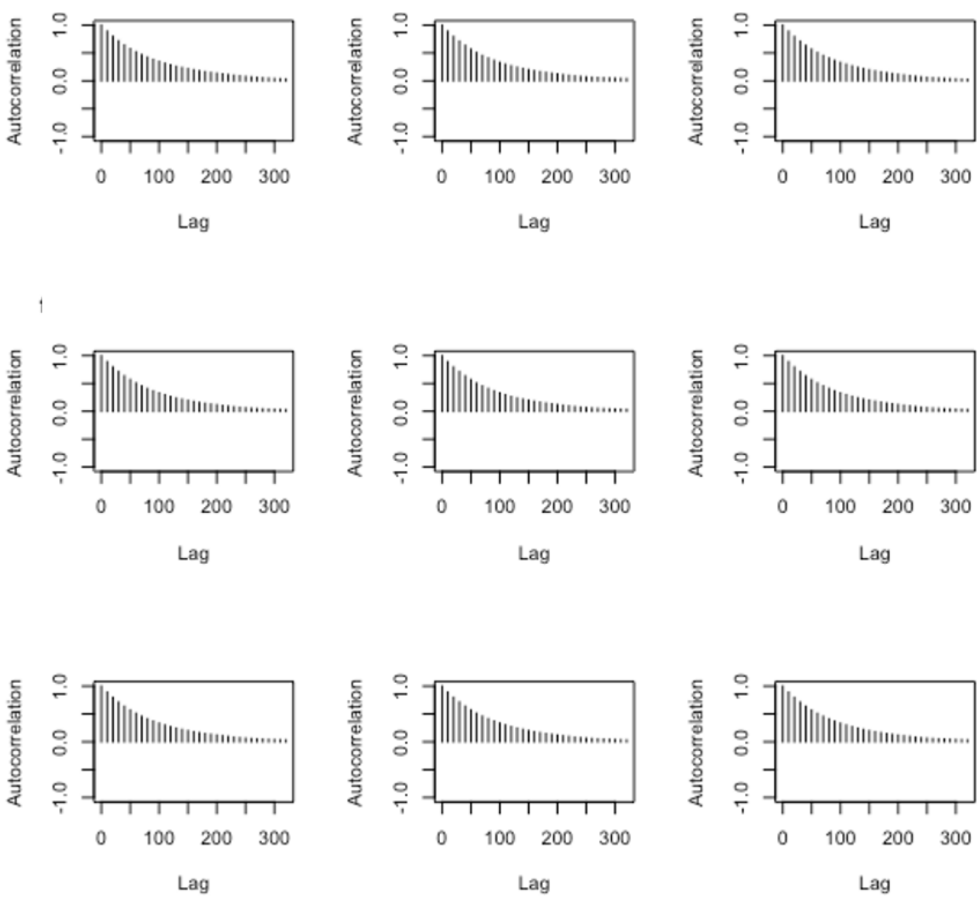


Figure 31. Sample of autocorrelations of regression coefficients for Model 2

Posterior Predictive Probabilities

The Bayesian Multinomial Model was fit for the exposure data and the diagnostics showed that the Metropolis-Hastings algorithm converged to the posterior distribution. The results of the predictive probabilities for Model 2 is summarized in

Table 11. The predicted gesture choice (i.e., the gesture with the highest probability) is provided for all combinations of functions and exposure where the exposure is either session 1, 2, or 3.

Table 11. Posterior predictive probabilities of intuitive gesture choice based on exposure and the contextual task for Model 2

Function [#]	Intuitive Gesture Mapped			
	Unspecified	**Session 1**	**Session 2**	**Session 3**
Go to the next slide [1]	Swipe hand left (0.296)	Swipe hand left (0.379)	Swipe hand left (0.317)	Swipe hand left (0.246)
Go to the previous slide [2]	Swipe hand right (0.220)	Swipe hand right (0.205)	Swipe hand right (0.214)	Swipe hand right (0.233)
Zoom in on the image [3]	Reverse full pinch (0.232)	Reverse full pinch (0.227)	Reverse full pinch (0.216)	Reverse full pinch (0.249)
Zoom out on the image [4]	Full pinch (0.211)	Full pinch (0.214)	Full pinch (0.191)	Full pinch (0.202)
Increase the volume [5]	Swipe hand up (0.155)	Swipe hand up (0.167)	Swipe hand up (0.125)	Swipe hand up (0.124)
Decrease the Volume [6]	Swipe hand down (0.149)	Swipe hand down (0.134)	Swipe hand down (0.178)	Swipe hand down (0.156)
Acknowledge the message [7]	Push fingers (0.318)	Push fingers (0.287)	Push fingers (0.345)	Push fingers (0.332)
Select "College of Bus." [8]	Push fingers (0.504)	Push fingers (0.470)	Push fingers (0.503)	Push fingers (0.524)
Undo the action [9]	Circle (0.425)	Circle (0.384)	Circle (0.490)	Circle (0.410)

Note: All values in the table represent the probability that y_{n+1} takes on a gesture, *k*, given the type of user

All combinations of exposure to the respective contextual task was mapped to the same gesture, consistent with the results from Chapter 4 in the frequency analysis.

Discussion

The objective of this chapter was to address the third research question of this dissertation: Q3 – How accurately can gestures be statistically predicted? Intuitive gesture choice is highly individualized thus making predictive analytics a difficult problem. Bayesian methods were used to model intuitive gesture choice because of the complexity of the multivariate data set. Both the expertise and exposure data were modelled, and posterior predictive probabilities were calculated for every combination of contextual task and user type as well as for every contextual task and session number. Intuitive gesture choice is multinomial data, and straightforward Bayesian analyses, such as using Multinomial-Dirichlet conjugate priors, are not capable of factoring in individual information such as expertise level or exposure level. Thus, two Bayesian multinomial logistic regression models were built using Metropolis-Hastings algorithms to model intuitive gesture choice based on expertise (Model 1) and exposure (Model 2).

The human-based gestural data was modelled successfully through a Bayesian approach to define posterior predictive probabilities for gesture choice. Overall, the results show that even when a human-based gesture development approach is taken and users are not forced to learn a predefined gestural language, it is possible to anticipate intuitive gesture choice depending on the contextual task and individual factors. Model 1 showed that when no individual factors are considered (i.e., user type not specified), the predictions are relatively non-informative. However, the predictions improve when specifying the expertise level. Model 2 showed that the predictions were consistent across exposure levels. The most notable predictions in Model 2 were for “Function 8 – Select College of Business from the Menu” and “Function 9 – Undo the action” as the predicted gesture choice was predicted with a probability around 0.50.

Both Model 1 and Model 2 had similar results with the frequency data from Chapters 3 and 4. The predictions were consistent with the intuitive gestures mapped via the frequency analysis. Future work should explore additional individual factors to increase the probability of predicted intuitive gesture choice, such as combining expertise and exposure into one analysis under one context as well as extending the exposure time to more than three experimental sessions.

There are some limitations with the Bayesian approach used in this chapter. Model 1 and Model 2 both used uninformative priors for the regression coefficients, and informative priors could have been used in the analysis. However, there was no previous data to generate an accurate prior distribution for the regression coefficients. Future work in the anesthesia context or general HCI context can utilize the results of this dissertation to build informative priors to be used in future Bayesian predictions for intuitive gesture choice. Furthermore, other covariates could be included such as trust and acceptance of the technology. Future work should also explore other sampling methods such as hybrid Metropolis-Hastings approaches with both independent and random-walk samplers as well as Gibbs sampling by deriving full conditional distributions. Furthermore, college students were used to some degree in both models, so generalizability of the predictions is confined to the population for the data. Future work should investigate more diverse populations in building the Bayesian models.

CHAPTER 6: USING A BOTTOM-UP APPROACH TO DEFINE AND CLASSIFY GESTURES TO IMPROVE GESTURE-FUNCTION MAPPINGS

Introduction

This chapter discusses ongoing work that supplements this dissertation. This work does not directly relate to a specific research question but overall adds value to the study of gestural control. The purpose behind this dissertation was to better understand the human factors issues related to gestural human-computer interaction, specifically the effect of exposure and expertise, and to understand how accurately gestures can be predicted within a particular context. There is currently a tradeoff between usability and accuracy from human-based and technology-based development approaches. This dissertation contributed to improving the human-based methodologies and understanding humans' gestural behavior. The results of this dissertation suggest that the human is an important piece to consider in gestural development.

Although the work in this dissertation took steps towards understanding the human side of gestural systems, there is still opportunity to further bridge the gap between the human-centered and technology-centered worlds in coming to a compromise between recognition accuracy and usability. If it is necessary to take a human-based approach, then the question remains of how recognition accuracy can be increased. One way that developers and programmers have combated the negative effects on recognition accuracy is by improving the recognition capabilities of the software, and there is considerable interest in the research community to develop methods which ensure a high recognition accuracy. One means is through deep neural networks which learn particular features for gesture recognition from the raw data from the camera (Huang et al., 2015). The deep neural network recognition approach has been shown to increase recognition accuracy to about 99% (Huang et al., 2015). Other methods have shown equal success such as hidden markov models, support vector machines, Eigenspace-based methods, and dynamic programming (Pisharady & Saerbeck, 2015).

Despite the progress in recognition software, there continues to be little research on the human factors side in reducing the impacts of the usability/accuracy tradeoff outside of this dissertation. In all human-based approaches, once the gesture-function mappings are defined then the mappings are programmed through the hand tracking algorithms. The camera reads XYZ coordinates of the hand gesture, and the XYZ coordinates are then translated into features of gestures (e.g., palm is facing camera, angles between fingers) then the features combine into one gesture – the original gesture-function mapping from user-elicitation studies. Thus, the gestures are recognized from an entirely bottom-up approach: starting with the raw XYZ data, to feature recognition, to identifying the entire gesture. Therefore, it may be advantageous to define a gesture by its features rather than as a holistic unit creating consistency with the capabilities of recognition software.

Despite the bottom-up characteristic of hand tracking algorithms, human-based approaches to eliciting and defining gestures take a rather top-down approach to classifying and describing the gestures. Researchers typically classify and describe the gesture as a single holistic unit in a top-down fashion (Aigner et al., 2012; Choi, 2012; Dong et al., 2015; Epps, Lichman, & Wu, 2006; Freeman et al., 2009; Henze, Löcken, Boll, Hesselmann, & Pielot, 2010; Höysniemi et al., 2004; Jacob et al., 2013; Kühnel et al., 2011; Mauney, Howarth, Wirtanen, & Capra, 2010; Morris et al., 2010; Pereira et al., 2015; Stern et al., 2006; Wobbrock et al., 2009). There are several disadvantages to the top-down approach of classifying gestures as a single unit. Gestures are highly individualized (Stern et al., 2008), so when classifying gestures as a unit and then analyzing for frequency of use under the consensus approach, the gesture-function mapping results may not show a clear consensus in the gestures, just a group of highly individualized gestures. It has also been shown that this individualization, specifically the interpretation of a gesture, is highly dependent on an individual's culture and past experiences (Mauney et al., 2010; Rautaray & Agrawal, 2015). Furthermore, it has been shown that both context (Ardito et al., 2014; Jacob & Wachs, 2014a; Jacob et al., 2013; Nielsen et al., 2004; Wigdor & Wixon, 2011) and domain expertise influence gesture-function mappings (Jurewicz & Neyens, 2017; Jurewicz et al., 2018). However, a top-down classification approach may not account for contextual, expertise, and other individual differences.

The alternative to the top-down, unit-based classification is to take a bottom-up approach and decompose a gesture into its features. If a gesture is broken down into lower level elements, there may be less cultural, context, and domain dependency as objective features of gestures (e.g., how many fingers are used) are independent from the semantics and the context of the application. Additionally, the features of gestures may translate better into recognition software since gesture recognition software uses the lowest level element of a gesture (e.g., XYZ coordinates) to calculate features (e.g., angles between fingers) of gestures (Huang et al., 2015). Some frameworks for designing midair gesture interfaces have been proposed that seek to define gestures by more than one feature such as Wobbrock, Morris, & Wilson's (2009) taxonomy of 2D, surface gestures which defined the gesture's form (e.g., static pose), nature, binding, and flow. Additionally, Uva et al. (2019) proposed a framework for 3D midair gestures which included end user analysis, gesture elicitation, vocabularies definition, and a validation procedure. These approaches are unique in that they seek to define a singular gesture in more than one way. However, there is little work investigating the advantages of bottom-up gestural classification based on a gesture's features compared to more traditional top-down approaches of unit classification. The overall objective of this study was to propose a strategy for decomposing gestures into features from a human-based gesture study and to compare the results of the bottom-up, feature extraction approach to the top-down, unit based approach. It is anticipated that the bottom-up classification approach would offer several advantages from both a technical and human factors perspective.

Methods

A top-down, unit-based taxonomy and a bottom-up, feature extraction taxonomy for classification were developed by the research team based on previous work (Jurewicz & Neyens, 2017; Jurewicz et al., 2018). The gestures classified by the taxonomies were from the novice data from Chapter 3. All gestures were classified via the top-down taxonomy and the bottom-up taxonomy. The two approaches were compared by the consensus of the gesture vocabulary sets as defined by Nielsen et al.'s (2004) method for identifying intuitive gesture-function mappings.

Top-Down Gesture Taxonomy

The top-down, unit based taxonomy was built to classify gestures as a unit according to the gesture's posture, pose, or movement direction. This classification method is similar to other approaches used in the literature (Choi, 2012; D. Freeman et al., 2009; Henze, Löcken, Boll, Hesselmann, & Pielot, 2010; Jacob et al., 2013; Pereira et al., 2015; Stern, Wachs, & Edan, 2006). The top-down gesture taxonomy consisted of a list of gesture names (e.g., "thumbs up") and operational definitions (e.g., "a static position of the thumb pointed up and all other fingers tucked into the palm"). The gesture dictionary was the same classification method used in Chapter 3-5.

Bottom-Up Gesture Taxonomy

The bottom-up, feature extraction taxonomy was built to classify the different possible features of gestures (Figure 32). The taxonomy consisted of a multilevel classification with 9 potential classifications for one gesture.

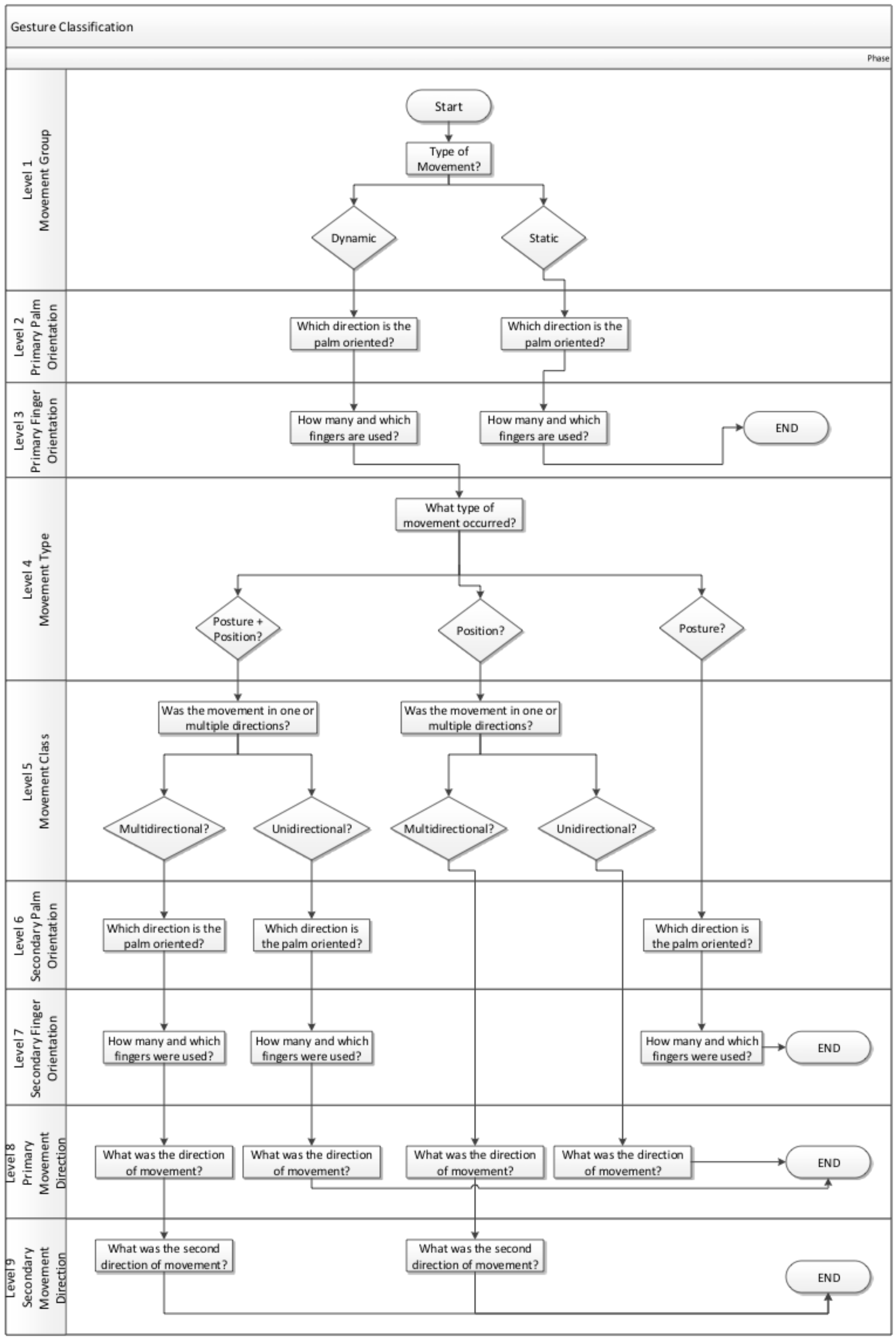


Figure 32. Bottom-up gesture classification taxonomy

The first feature of a gesture analyzed was the movement group (Level 1 in Figure 32), and gestures were classified as either static or dynamic. Gestures labelled as static were then broken down into the primary palm orientation feature (i.e., Level 2 in Figure 32, which direction is the palm facing) and primary finger orientation feature (i.e., Level 3 in Figure 32, how many and which fingers were used in the gesture). There were no other features extracted for static gestures.

After identifying the movement group, dynamic gestures were classified for primary palm and finger orientation features. Then, dynamic gestures were broken down into the movement type feature (Level 4 in Figure 32). The movement types could either be a posture change, a position change, or a posture and position change. For gestures with posture changes, the additional features extracted were the secondary palm orientation feature (Level 6 in Figure 32, which direction is the palm facing in the second part of the gesture) and the secondary finger orientation feature (Level 7 in Figure 32, how many and which fingers were used in the second part of the gesture).

For gestures with position changes, the next feature extracted was movement class. The movement class (Level 5 in Figure 32) could either be unidirectional, where the movement occurred in only one direction, or multidirectional, where multiple position changes occurred (e.g., waving the hand). Therefore, the additional features extracted for position changes were the primary movement direction (Level 8 in Figure 32, which direction did the hand move) and if applicable, the secondary movement direction for multidirectional gestures (Level 9 in Figure 32). As for the last dynamic gesture movement type, position and posture change, the additional features extracted were secondary palm orientation and secondary finger orientation, movement class, primary movement direction, and if applicable, secondary movement direction.

Results

Top-Down Gesture Classification

The gesture-function mappings from the top-down approach are summarized in Table 12. This table shows the intuitive gesture mapped as well as the gesture that was performed second most frequently. These findings are described in detail in Chapter 3.

Table 12. Gesture-function mappings generated from the top-down classification approach

Function [#]	Intuitive Gesture Mapped (n)	2nd top gesture (n)
Start the flow [1]	Thumbs up (29)	Swipe hand up (14)
Stop the flow [2]	Five up (17)	Push hand (14)
Inc. the flow [3]	Swipe hand up (33)	Thumbs up (16)
Dec. the flow [4]	Swipe hand down (33)	Thumbs down (17)
Silence alarm [5]	Swipe hand left (15)	Swipe hand right (16)
Ack. the message [6]	Thumbs up (46)	Okay (16)
Heart rate normal? [7]	Thumbs up (54)	Okay (7)
Pulse ox normal? [8]	Thumbs up (25)	Thumbs down (23)
Select heart rate [9]	Push finger (25)	Three up (19)
Cancel the request [10]	Swipe hand left (25)	Thumbs down (13)

Bottom-Up Gesture Classification

For feature extraction, it was important that all gestures were from the perspective of the same hand, whether left or right handed gestures, in order to ensure consistency in the feature mappings. Only three participants did the experiment with their right hand, so this data was transformed into left handed data so that the analysis would be consistent for palm orientations. For example, when right-handed gestures had a palm orientation of "left," this mirrored a left-handed gesture with a palm orientation of "right;" therefore, the right handed palm orientation would translate to "right."

The feature extraction results are summarized in Table 13. All functions mapped to intuitive features. "Select heart rate" (Function 9) was the only function where it was unclear if there was an intuitive feature mapping in the "gesture type" feature. Regardless, other features for Function 9 had intuitive mappings. Three functions ("acknowledge the message", "is heart rate normal?", "is pulse oximeter normal?") were mapped to static gestures, so these functions did not map further to features related to movement. The rest of the gesture mappings were dynamic, but none were categorized as posture changes or multiple position changes; therefore, no functions were mapped to features related to a secondary posture or secondary movement direction. The finger orientations are listed as a five-digit number representing the how many and which fingers were used, as shown in Figure 33. The order of digits represents the thumb, index, middle, ring, and pinky finger, a 0 corresponds to that finger not being extended, and a 1 means the finger was extended in the function. For example, "thumbs up" would classify as "10000" as only the thumb is activated in the gesture (See Figure 33).

Table 13. Gesture-function mappings generated from the bottom-up classification approach

	Function									
	1 (n)	2 (n)	3 (n)	4 (n)	5 (n)	6 (n)	7 (n)	8 (n)	9 (n)	10 (n)
Level 1 – Gesture Type	Dynamic (49)	Dynamic (53)	Dynamic (67)	Dynamic (60)	Dynamic (66)	Static (72)	Static (75)	Static (74)	Static (46)/Dynamic (44)	Dynamic (59)
Level 2 – Primary Palm Orientation	Right (38)	Forward (64)	Forward (41)	Forward (36)	Forward (56)	Right (49)	Right (55)	Forward (34)	Forward (57)	Forward (47)
Level 3 – Primary Finger Orientation	10000 (36)	11111 (60)	11111 (44)	11111 (46)	11111 (69)	10000 (48)	10000 (56)	10000 (51)	01000 (29)/ 01110 (19)	11111 (60)
Level 4 – Movement Type	Position (47)	Position (47)	Position (63)	Position (57)	Position (55)				None (46)/ Position (34)	Position (52)
Level 5 – Movement Class	Single (47)	Single (44)	Single (63)	Single (57)	Single (46)				Single (34)	Single (46)
Level 6 – Secondary Palm Orientation										
Level 7 – Secondary Finger Orientation										
Level 8 – Primary Movement Direction	Up (26)	Forward (19)	Up (50)	Down (43)	Left (26)				Forward (28)	Left (30)
Level 9 – Secondary Movement Direction										

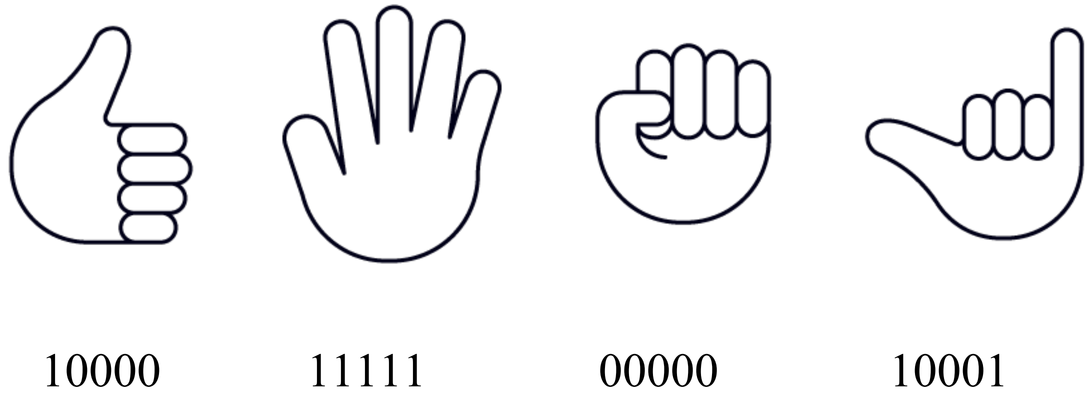


Figure 33. Examples of five digit number translations for a gesture's finger orientation feature

Comparison of Approaches

The results for “decrease the flow of anesthesia gas [2]”, “increase the flow of anesthesia gas [3]”, “acknowledge the message [4]”, “is heart rate normal? [7]”, and “cancel the request [10]” are consistent between the top-down and bottom-up classification approaches in that the features extracted in the bottom-up approach match the gesture mapped in the top-down approach. For example, “is heart rate normal?” was mapped to three features: static, palm orientation to the right, and a finger orientation of 10000 (see Figure 33). These features reflect the features of “thumbs up,” which was the intuitive mapping for “is heart rate normal?”.

There were four functions which did not have clear gesture-function mappings from the top-down approach as shown by the similarity in frequency between the top two performed gestures. These were “start the flow of anesthesia gas [1]”, “stop the flow of anesthesia gas [2]”, “silence the alarm [5]”, and “select heart rate [9].” However, the bottom-up classification approach showed that these functions had intuitive features. For example, with “increase the flow of anesthesia gas [3]”, there appears to be a consensus that the intuitive features are dynamic and single position movement with the movement direction as up. The participants may have differed in terms of posture but agreed on a dynamic upward movement.

Discussion

The objective of this study was to propose a strategy for decomposing gestures into features from a human-based gesture study via a bottom-up approach and to compare the results to the gesture-function mappings from a traditional top-down approach. The results of this study showed that there appears to be consensus across the group for particular features despite the lack of agreement across participants for one intuitive gesture. The bottom-up approach uncovers agreements that are not present in the top-down approach. Thus, the bottom-up approach proposed can maintain the intuitive benefits associated with human-centered gesture classification while also supporting bottom-up driven recognition accuracy.

Utilizing intuitive features has several design advantages including facilitating the use of having multiple gestures that have the same underlying features for one function, avoiding overlapping gestures across functions, and removing the dependency of mappings to semantics, culture, or past experiences.

In the feature extraction approach, there is a group of intuitive features which can potentially be used in the design of a gesture or set of gestures. For example, with the function "silence the alarm" the top gestures were swipe hand left and swipe hand right, but the features extracted were dynamic, forward, all fingers used, and a position change in one direction. A gestural interface designer could take these features and map the function to a gesture that is a dynamic movement of the hand in one direction, whether that direction is left or right, and the users could then have some flexibility in performing a gesture. Therefore, feature extraction classification approach facilitates the use of multiple gesture-function mappings.

Three functions in this study were mapped to "thumbs up" under the top-down approach. When a single gesture is mapped to multiple functions, the gestural system needs to be capable of understanding the context in which the gesture was performed to know which function is being done (Nielsen et al., 2004; Pereira et al., 2015). If the results from the top-down classification approach were the final gesture-function mappings, context sensitivity would be a concern in the gestural system with overlapping mapping. Using the feature extraction approach potentially avoids this issue. With the group of features extracted, gestural interface designers can take specific features to design gestures that do not overlap across tasks. For “start the flow of anesthesia gas,” the intuitive gesture mapped in the top-down approach was “thumbs up,” but under the feature extraction approach, the features were inconsistent with the static gesture of “thumbs up.” The features were actually more consistent with a dynamic gesture of “thumbs up” as in a swiping motion with a thumbs up posture where the position change was in the upwards direction. Therefore, under the feature extraction approach, gestural interface designers have the opportunity to utilize the group of features to design gestures that do not overlap between functions.

Looking specifically at the “Select heart rate” function, this display had four buttons and the third button down from the top was listed as “Heart Rate.” One of the top gestures under the unit-based approach was “push fingers” where the user pushed their finger towards the screen as if physically pushing that button. The other top gesture under the top-down approach was “three up” which was the pointer finger, middle finger, and ring finger pointed up and the rest of the fingers closed into the palm, as if the user was telling the system to choose the third option. In some context and cultures, choosing the number “three” would resemble the middle, ring, and pinky fingers pointed up. This gesture would classify as “okay” and would not capture the meaning of “three” under the top-down approach. However, both the “okay” and “three up” gestures would be classified as static gestures using three fingers under the bottom-up approach. This classification avoids any semantics that are dependent on the user’s culture as the mapping could be static with three fingers activated. Gestures, as a unit, are dependent on an individual’s past experiences and culture (Mauney et al., 2010; Rautaray & Agrawal, 2015); however, there is no evidence in the literature suggesting that features of gestures are dependent upon a culture or past experiences.

In addition to the several design advantages of the bottom-up approach, the features of gestures may translate better into system integration since gesture recognition software uses the lowest level element of a gesture (e.g., XYZ coordinates) to calculate features (e.g., angles between fingers) of gestures (Huang et al., 2015). Gestural systems should ideally have high accuracy and high usability. The bottom-up approach demonstrated in the current study maintains the intuitive benefits associated with human-centered gesture classification. Additionally, the bottom-up classification does not conflict with the bottom-up hand tracking algorithms as features as extracted instead of one gesture. Future work in gesture development research should continue to develop methods that yield both highly accurate and highly usable systems in order to guarantee reliability and overall success of the gestural system.

The taxonomy in the current study was able to capture features for the set of gestures performed in the experiment, but there are potentially additional features of gestures which were not captured in this study. Future work should fine tune and expand the classification of gestural features so as to be generalized to all potential gestures. Additionally, the feature extraction classification approach only accounted for single hand gestures since only single hand gestures were performed in the experiment;

CHAPTER 7: CONCLUSIONS

The overall objective of this dissertation was to understand and model how humans behave in 3D gestural input systems. Specifically, this dissertation investigated the influence of the anesthesia context, domain expertise, and exposure on gesture behavior. This work has shown the importance of considering individual factors when developing 3D gestural input systems such as domain expertise and exposure. This work also investigated a Bayesian approach to modeling human behavior, specifically for predicting intuitive gesture choice based on contextual task, expertise, and exposure. The Bayesian models showed that even when taking a human-based development approach, intuitive gesture choice can be predicted based off of expected natural responses and certain individual factors.

Limitations

This dissertation is a small step towards gestural systems which are highly accurate and highly usable, and there are some limitations in this work. College students were the primary population used to recruit participants, thus the results may not generalize to a larger, more diverse population. In investigating the effect of exposure, the participants completed the experiment three times with about 48 hours in between each session. Behavioral adaptations may continue to evolve over time, and three times may not be sufficient enough to investigate the effect of exposure. Furthermore, behavior may adapt as the entire work system changes such as technological changes in the anesthesia workstation, so system vulnerabilities and reliance to gestural control needs to be monitored for long-term use. Additionally, expertise was studied for domain novices and domain experts, and there may be further differences between novices and experts within a domain. All studies incorporated a Wizard-of-Oz approach to eliciting gestures, and it was necessary to do so to allow the participant flexibility in choosing gestures without the influence of technological limitations.

Future Work

Future work of this dissertation includes continuing to explore the individual factors which influence gesture behavior, especially within specific contexts such as anesthetic care in the OR. Future work should investigate other levels of expertise within anesthesia, and other domains, and over an extended period of time. The sample population should also extend beyond college students in order to ensure interpretations and predictions are reflective of actual system dynamics. Future work should also investigate how the 3D gestural technology actually integrates into the work system and explore variables such as user frustrations, reliance, and trust when the Wizard-of-Oz component is removed. Future data collection should continue to be tested in a Bayesian model for identifying accurate intuitive gesture choice based on factors which influence gesture behavior.

Gestural input technology can potentially be very impactful for the healthcare industry, especially anesthetic care. It is already known that there is widespread bacterial contamination in the anesthesia workstation (Birnbach et al., 2015; Loftus et al., 2011) but current anesthetic work does not support the addition of more hand hygiene steps (Jurewicz et al., Under Review). 3D gestural control in anesthesia is novel and innovative, and future work in this area should see how human behavior adapts when fully implemented into the work system. There are often multiple anesthesia providers present in the anesthesia workstation, and more providers may impact gesture behavior. There also may be safety critical concerns relative to particular functions in which gestural technology may not be appropriate. The variables explored in this dissertation, as well as other individual factors, should be studying in a larger anesthesia study.

Overall, this work provides insights into how natural user interfaces (NUIs) may be designed in the form of 3D gestural HCI. Gestures are already a natural form of communication, so gestures could be used for HCI instead of traditional input devices such as keyboard and mouse. There exists little work from the human factors side in investigating gestural displays despite the fact that 3D gestural systems already exist in homes and cars. NUIs and 3D gestural displays could potentially transform how humans interact with machines and technology and could be impactful for not just anesthesia but the healthcare industry at large as well as other complex human-machine systems.

This dissertation provides insights into how Bayesian statistics can be utilized in human factors research. There is little work in human factors and HCI in incorporating Bayesian methods despite the many advantages to using a Bayesian approach. Human-based data is complicated, messy, and many assumptions and translations of the data have to be made just to perform traditional frequentist statistics whereas a Bayesian approach embraces the mess of human data. This dissertation work shows how real data can be modelled without making assumptions or transformations and provides insights into how Bayesian statistics can be used more broadly in human factors applications. There are limitations to using a Bayesian approach, especially in terms of computation. The data in this dissertation were multivariate with multiple covariates, thus the computations of the Bayesian models are highly expensive to run in terms of computational power and time. Future work should develop packages and functions in R that facilitate efficient coding practices. Human-machine systems will continue to grow in complexity especially with the rapid release of new and innovative technologies, and it will be important to ensure that the modeling and design of the systems are reflective of true human-machine dynamics potentially through Bayesian modeling.

Broader Impacts and Intellectual Merit

Computers are an integral part of people's lives; however, traditional input devices such as keyboard and mouse still dominate the technical market. This dissertation sought to explore 3D gestural control and supported the investigation and understanding of gesture behavior. The future of human-computer interaction and human-machine systems is to have new and natural interaction methods. Leaders and pioneers in HCI, such as Bill Gates, believe in the future of natural user interfaces (NUIs):

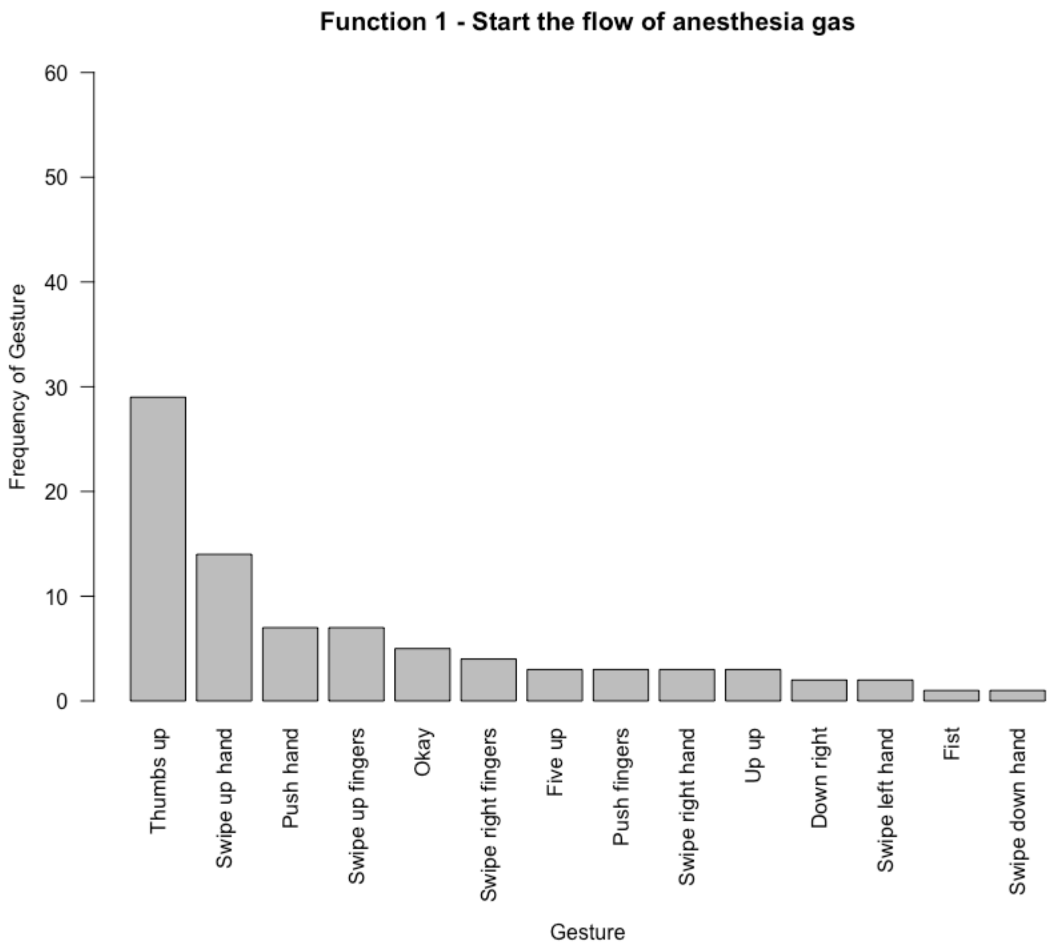
> "Until now, we have always had to adapt to the limits of technology and conform the way we work with computers to a set of arbitrary conventions and procedures. With NUI, computing devices will adapt to our needs and preferences for the first time and humans will begin to use technology in whatever way is most comfortable and natural for us." (Mortensen, 2017)

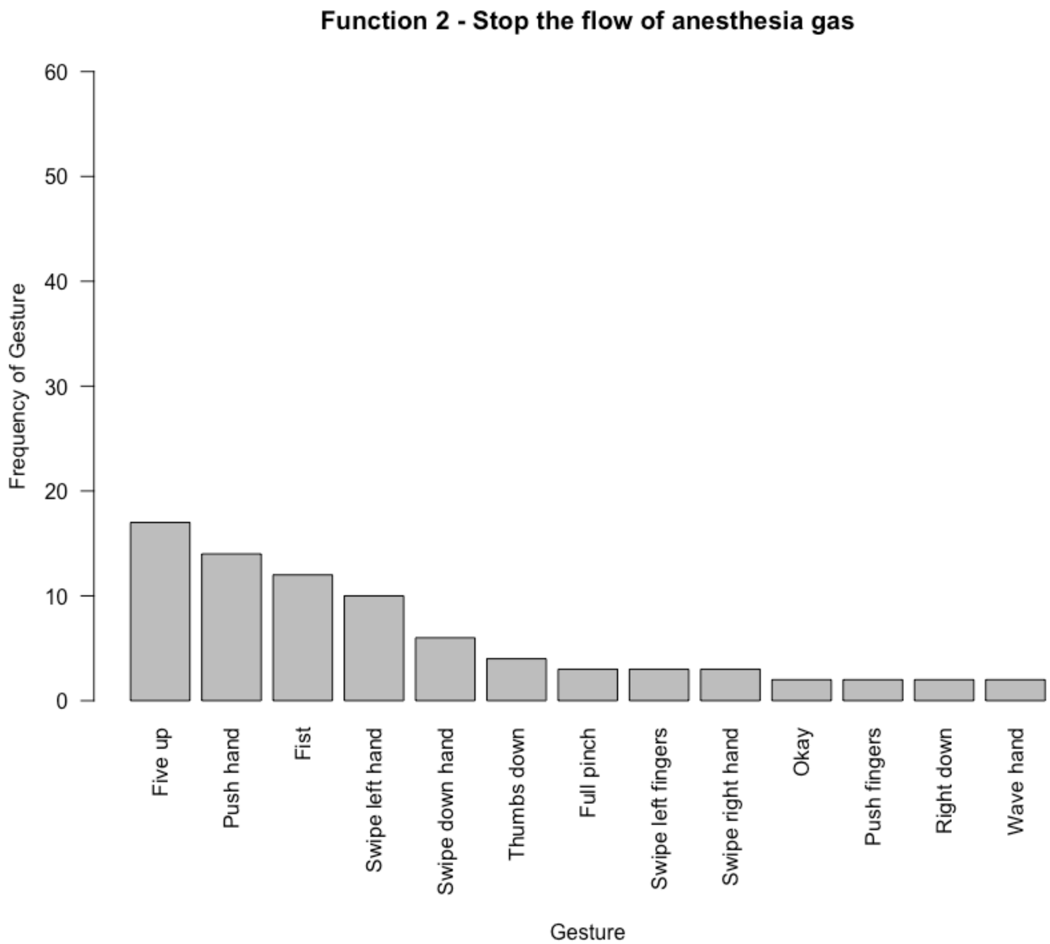
The current reality of the development of NUIs, specifically 3D gestural interfaces, is shown in Figure 2 where there are technology-based systems with high accuracy but low usability, or there are human-based systems that are highly usable but have low recognition accuracy. It is important for the future of 3D gestural interfaces, and NUIs broadly, to alleviate the accuracy/usability tradeoff so that systems are highly accurate and highly usable to ensure long-term reliability and use the system. HCI researchers, both on the technical and human sides, need to continue future work in refining methods to develop reliable NUIs. Human factors engineers and researchers need to continue to understand how humans behave in a gestural system and to identify advanced ways to model real human behavior. If gesture behavior can be explained and predicted, then motion tracking and gesture recognition software can be improved by including predictive analytics based on human data. The system may further be improved by incorporating human-based methods which are consistent with hand tracking algorithms, such as the novel bottom-up approach to classification presented in this dissertation. Eventually, adaptive 3D gestural interfaces can be developed that are based on the predictive analytics of intuitive gestural features of multiple user groups and across contexts and applications.

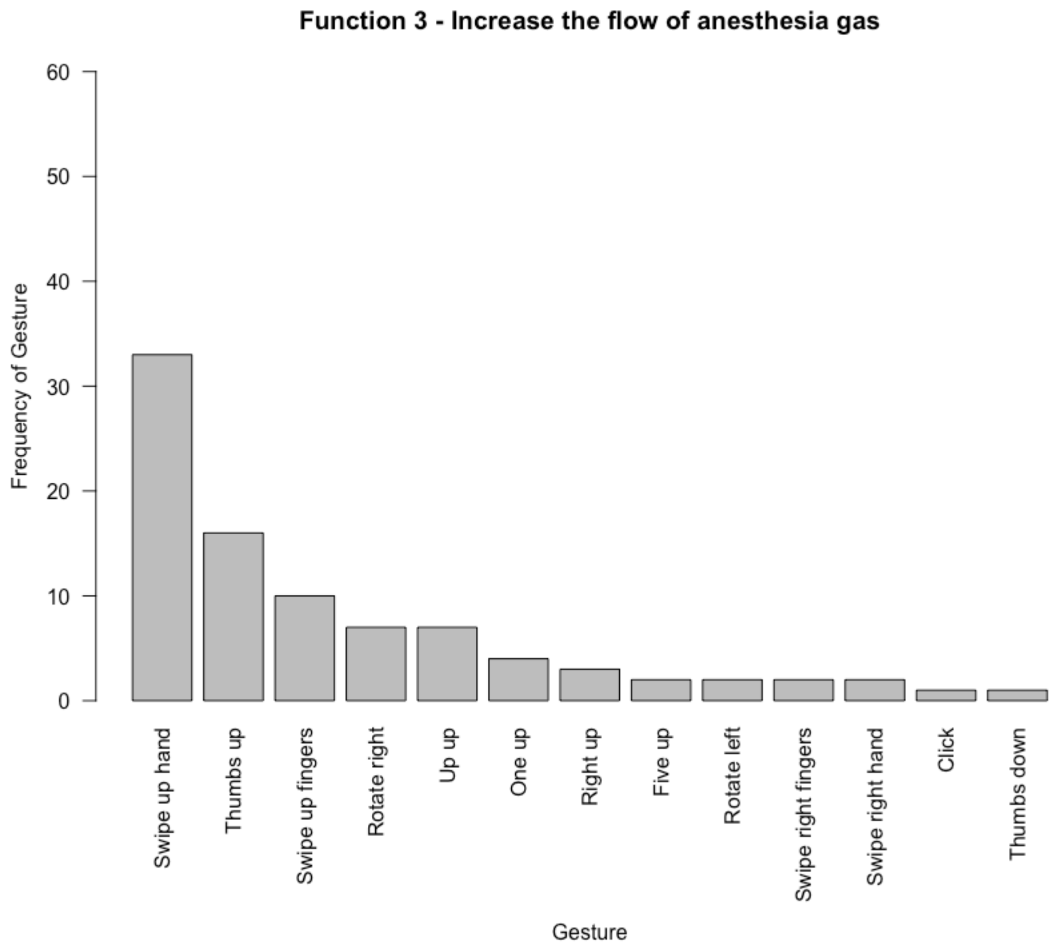
The human-technology frontier is dynamic, and the methodology demonstrated as well as the findings within this dissertation can be adapted to larger human-machine systems that include more than just 3D gestural control. The future of human-machine systems will incorporate more natural and comfortable means of interaction, and development of such systems will require an understanding of human behavior (e.g., Chapters 3 and 4 - the factors that influence gesture behavior), useful translations of human behavior to computer language (e.g., Chapter 6 – feature extraction, bottom-up approaches vs. higher level, top-down approaches to analyzing human behavior), and advanced statistical computations (e.g., Chapter 5 – Bayesian predictive analytics).

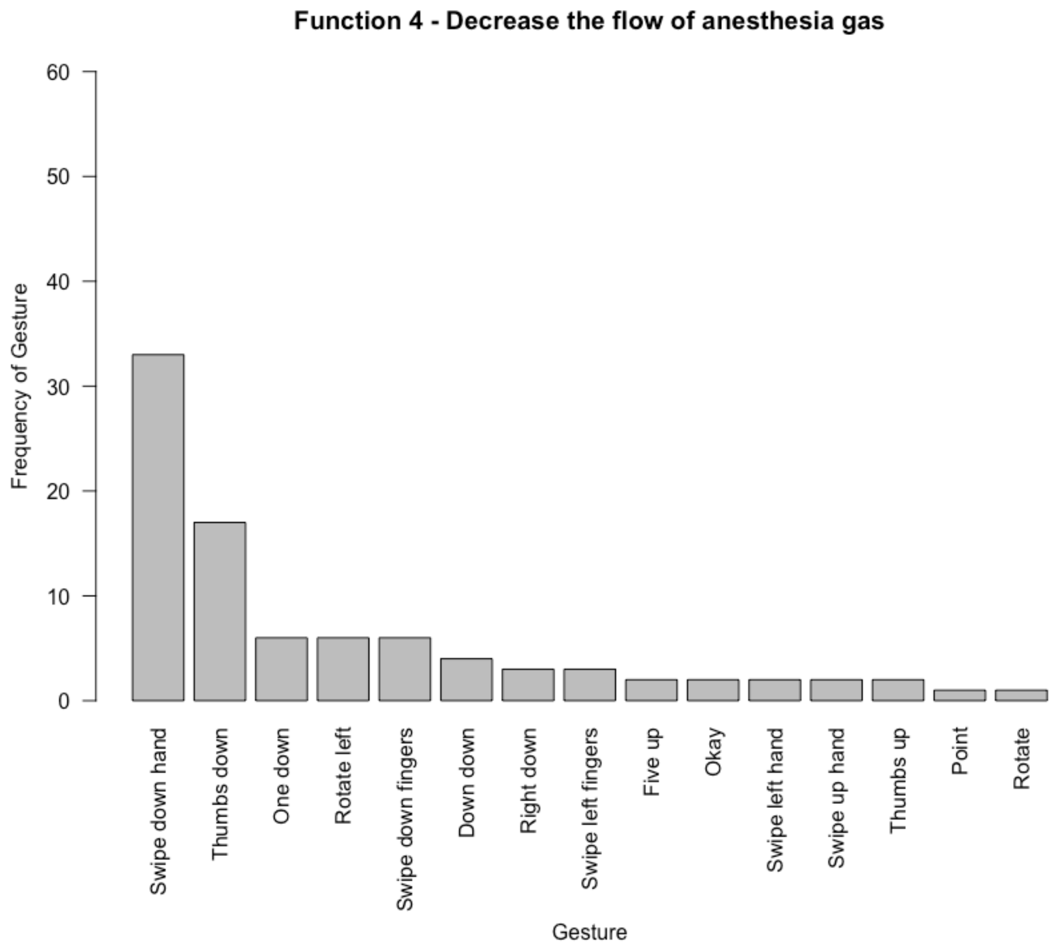
APPENDICES

Appendix A – Domain novices bar graphs for intuitive gestures in anesthesia

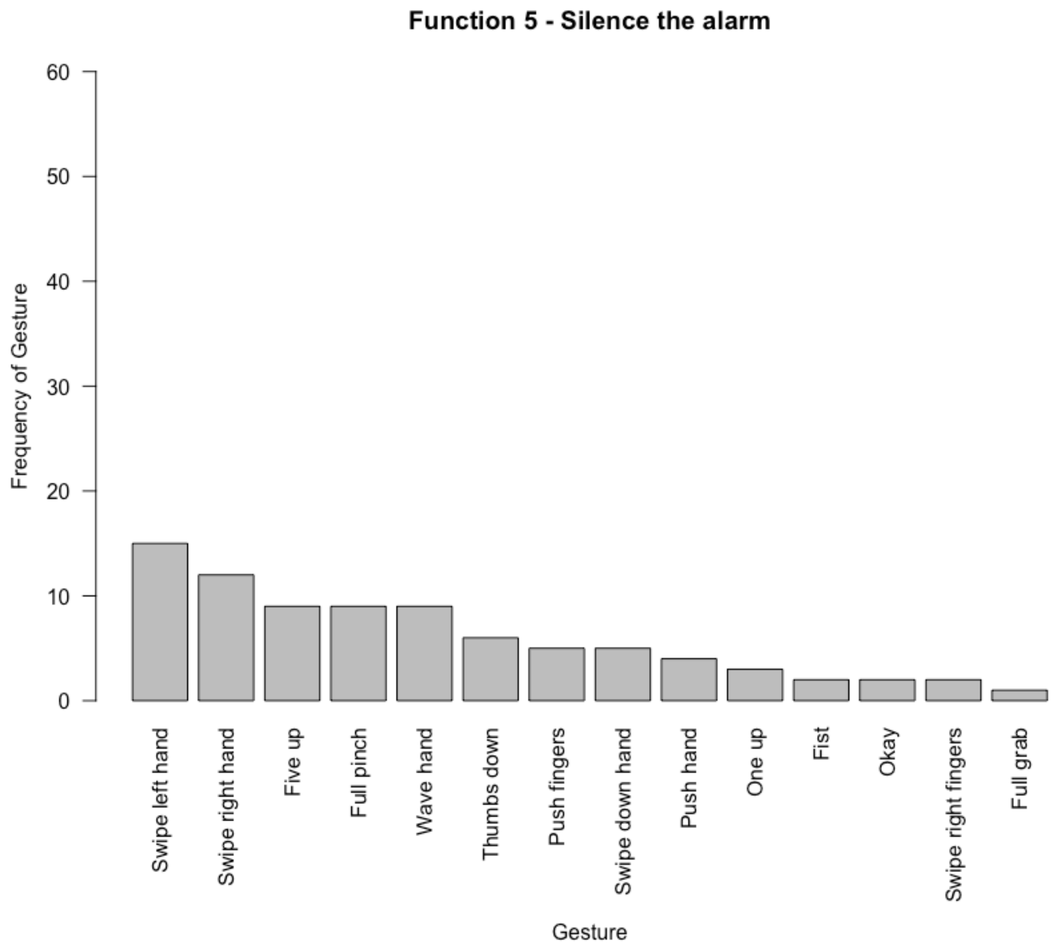


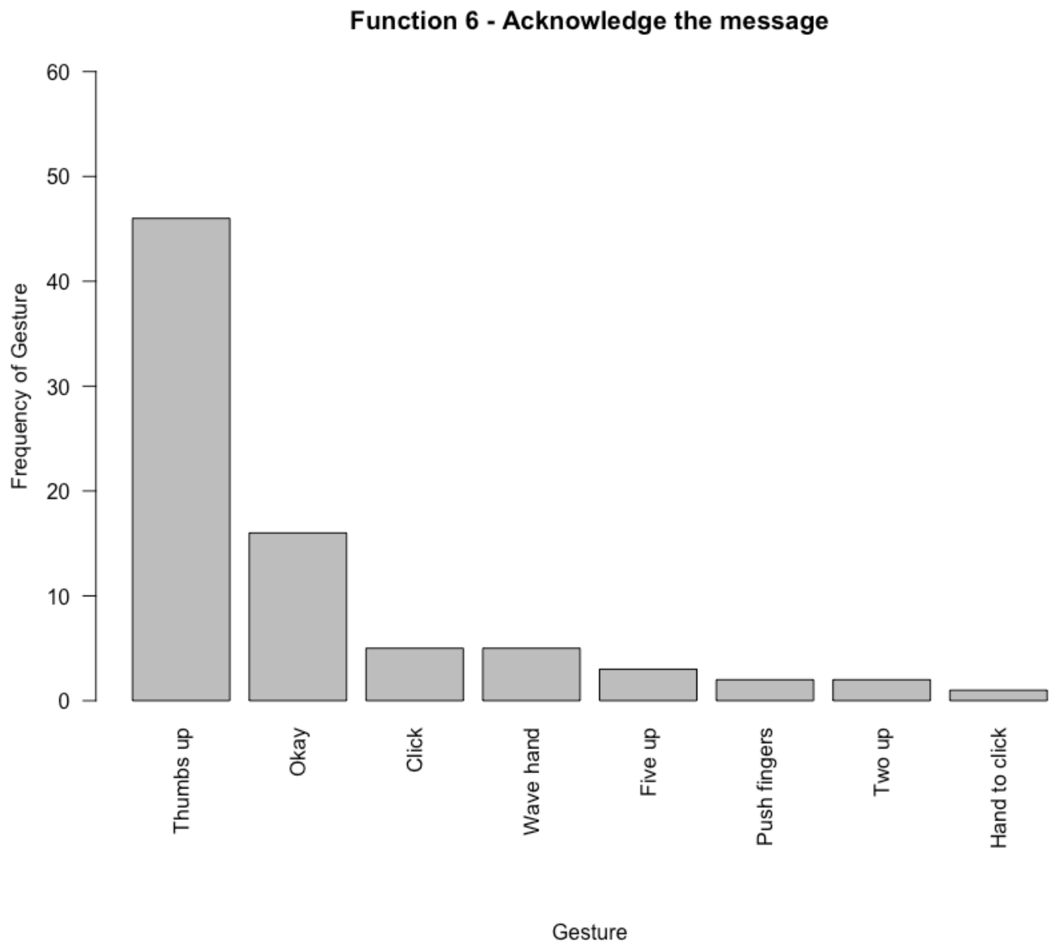

Function 2 - Stop the flow of anesthesia gas
Frequency of Gesture
60
50
40
30
20
10
0
Five up
Push hand
Fist
Swipe left hand
Swipe down hand
Thumbs down
Full pinch
Swipe left fingers
Swipe right hand
Okay
Push fingers
Right down
Wave hand
Gesture

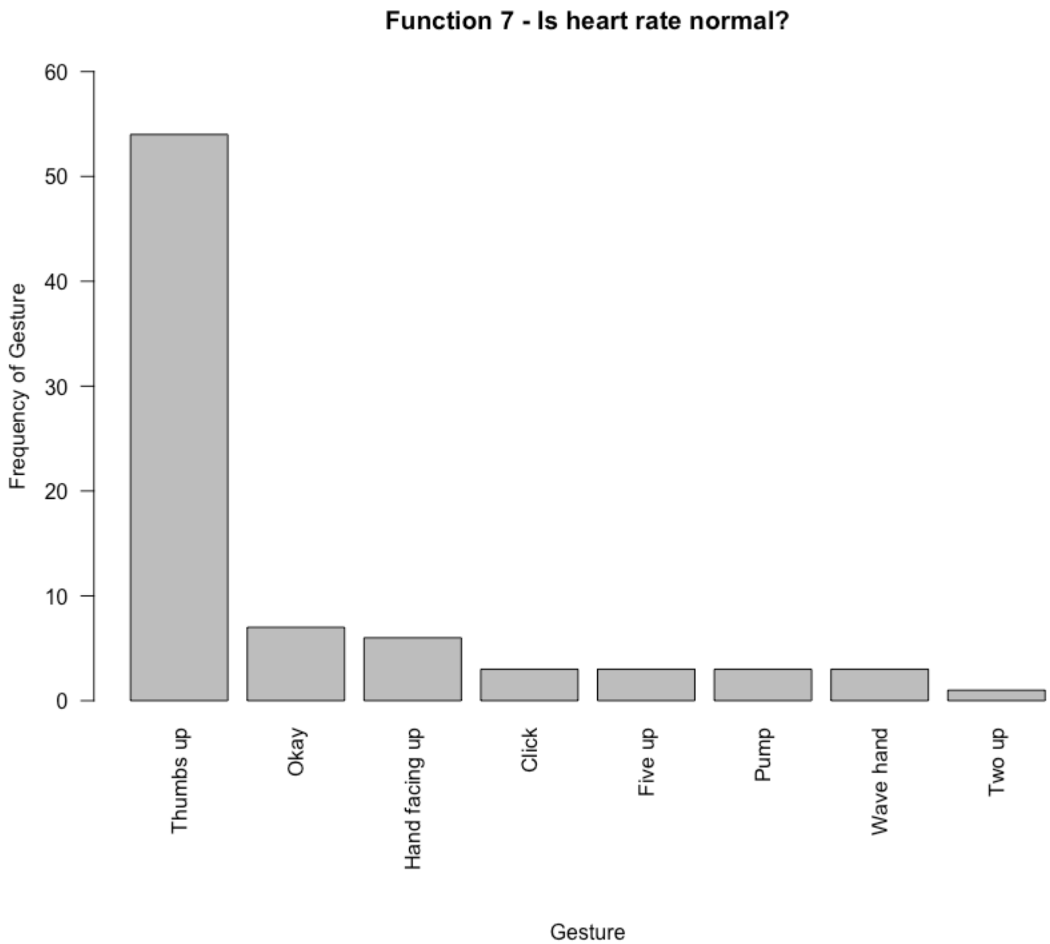

Function 3 - Increase the flow of anesthesia gas
Frequency of Gesture
60
50
40
30
20
10
0
Swipe up hand
Thumbs up
Swipe up fingers
Rotate right
Up up
One up
Right up
Five up
Rotate left
Swipe right fingers
Swipe right hand
Click
Thumbs down
Gesture



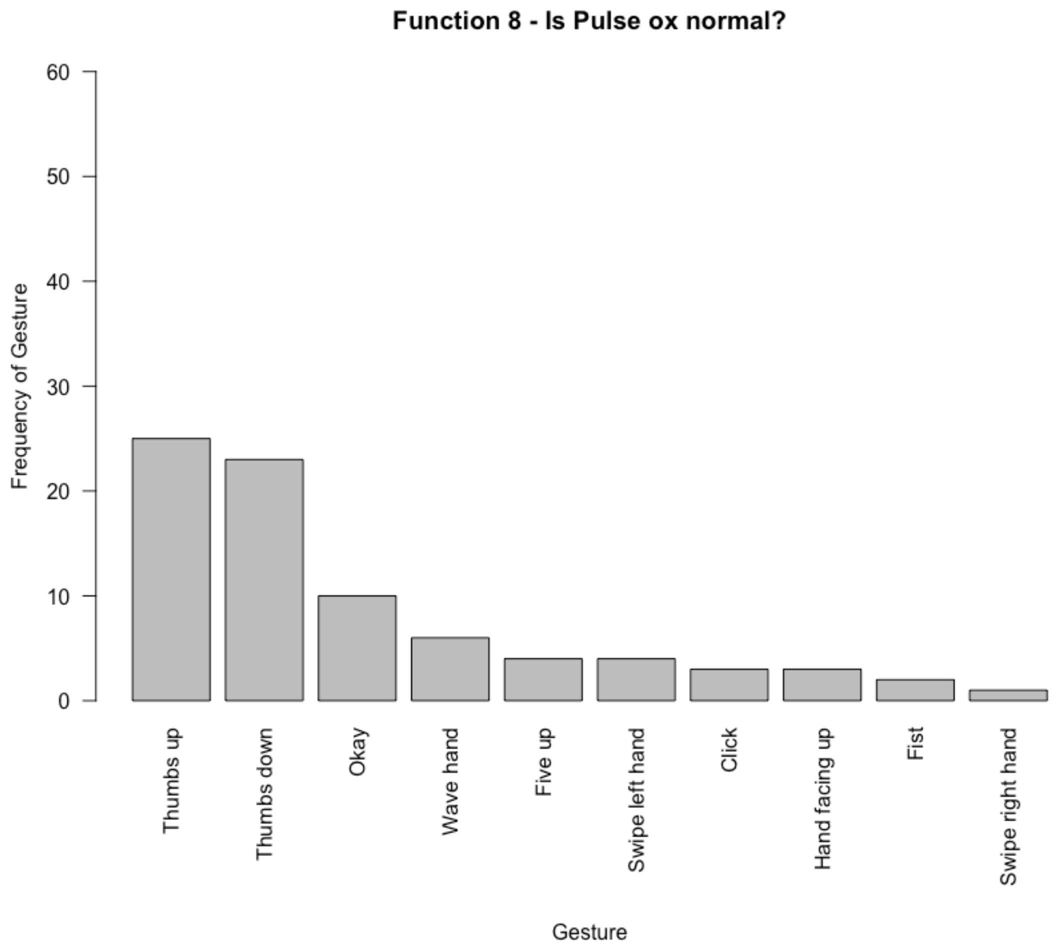
Function 4 - Decrease the flow of anesthesia gas
Frequency of Gesture
0
10
20
30
40
50
60
Swipe down hand
Thumbs down
One down
Rotate left
Swipe down fingers
Down down
Right down
Swipe left fingers
Five up
Okay
Swipe left hand
Swipe up hand
Thumbs up
Point
Rotate
Gesture


Function 5 - Silence the alarm
Frequency of Gesture
60
50
40
30
20
10
0
Swipe left hand
Swipe right hand
Five up
Full pinch
Wave hand
Thumbs down
Push fingers
Swipe down hand
Push hand
One up
Fist
Okay
Swipe right fingers
Full grab
Gesture

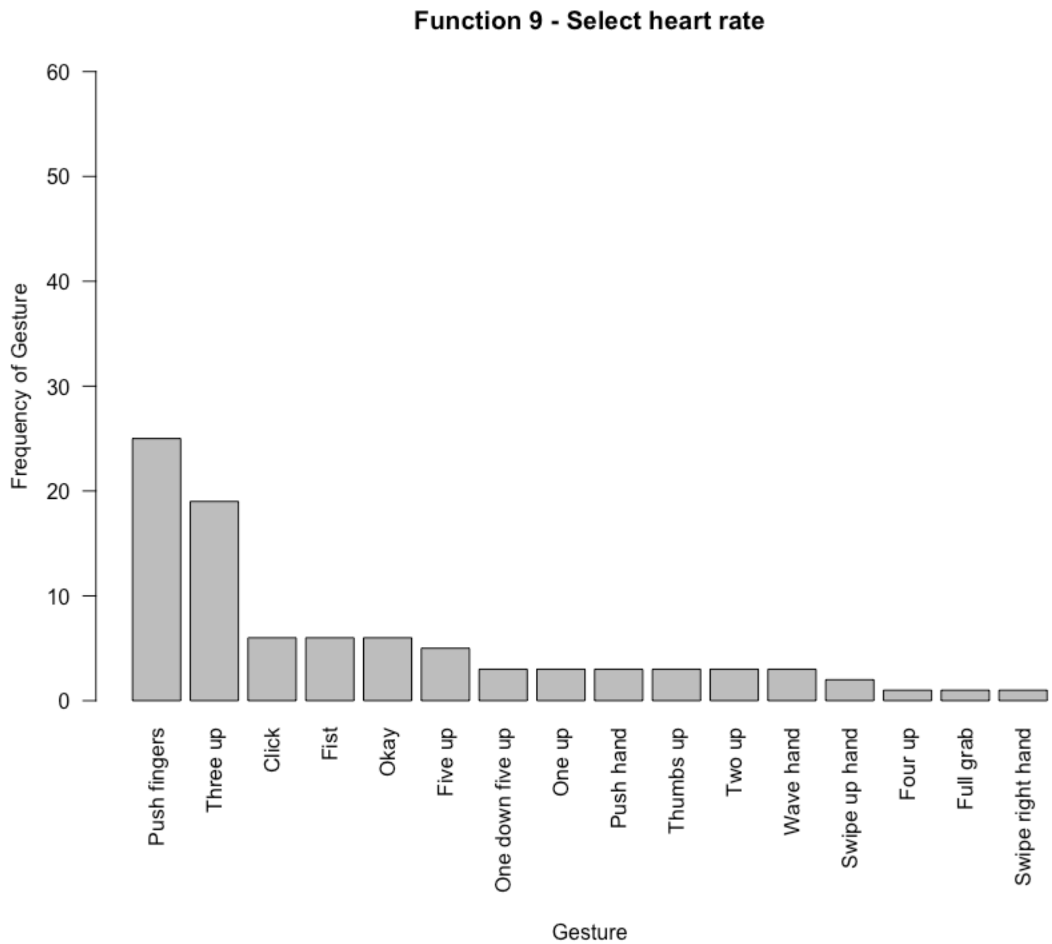

Function 6 - Acknowledge the message
Frequency of Gesture
60
50
40
30
20
10
0
Thumbs up
Okay
Click
Wave hand
Five up
Push fingers
Two up
Hand to click
Gesture

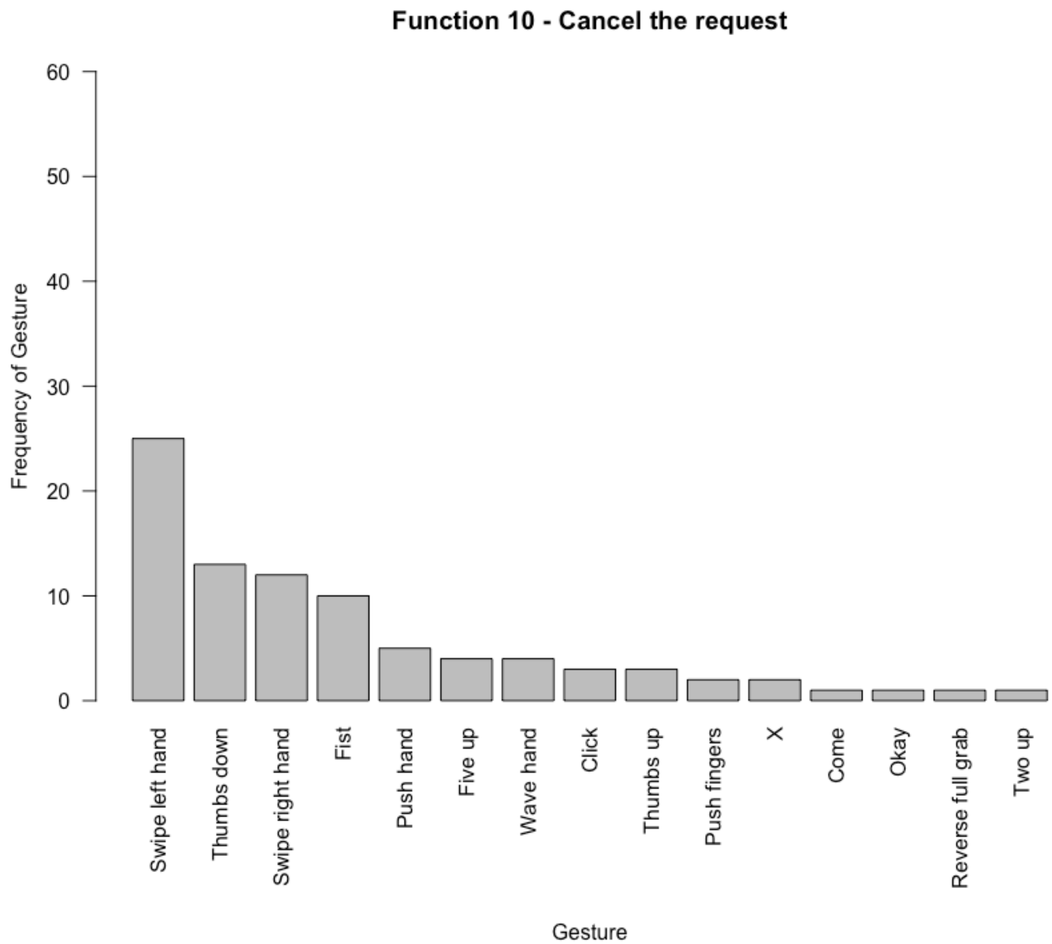


Function 7 - Is heart rate normal?
60
50
40
30
20
10
0
Frequency of Gesture
Thumbs up
Okay
Hand facing up
Click
Five up
Pump
Wave hand
Two up
Gesture



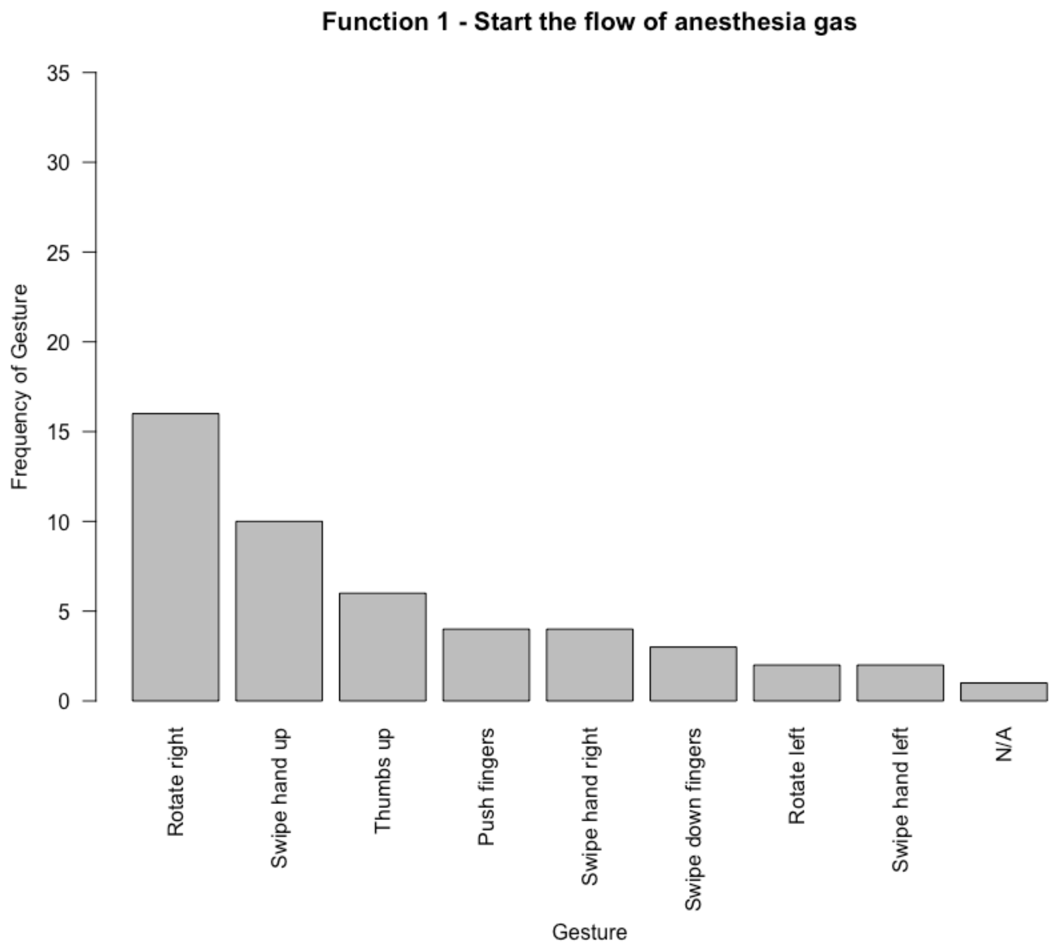
Function 8 - Is Pulse ox normal?
Frequency of Gesture
60
50
40
30
20
10
0
Thumbs up
Thumbs down
Okay
Wave hand
Five up
Swipe left hand
Click
Hand facing up
Fist
Swipe right hand
Gesture

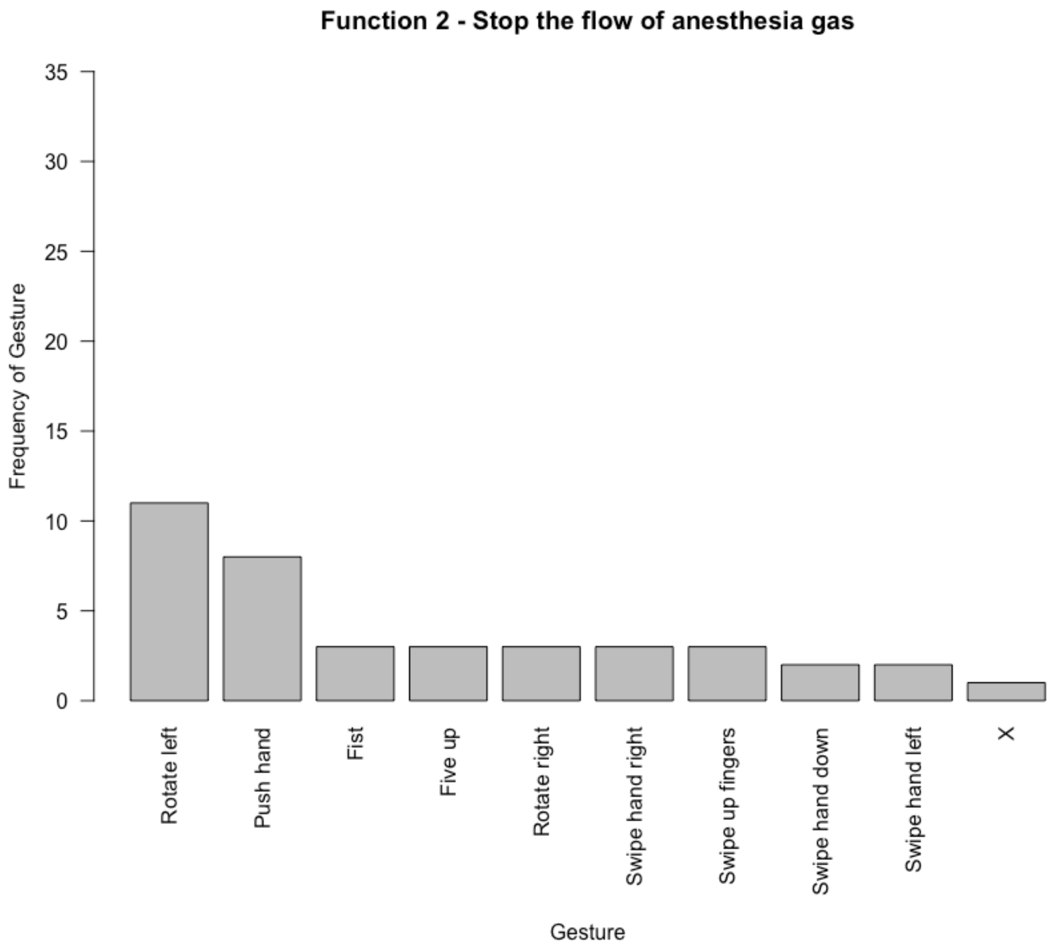

Function 9 - Select heart rate
Frequency of Gesture
60
50
40
30
20
10
0
Push fingers
Three up
Click
Fist
Okay
Five up
One down five up
One up
Push hand
Thumbs up
Two up
Wave hand
Swipe up hand
Four up
Full grab
Swipe right hand
Gesture

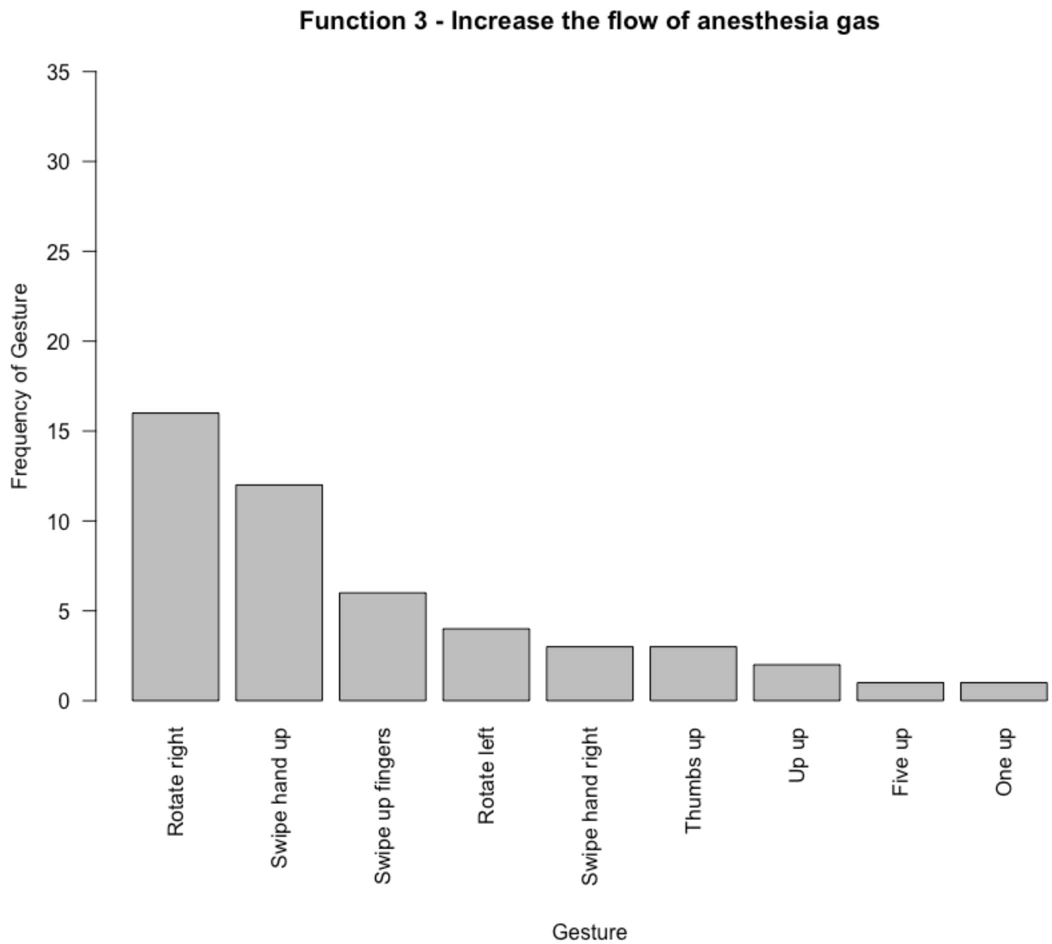


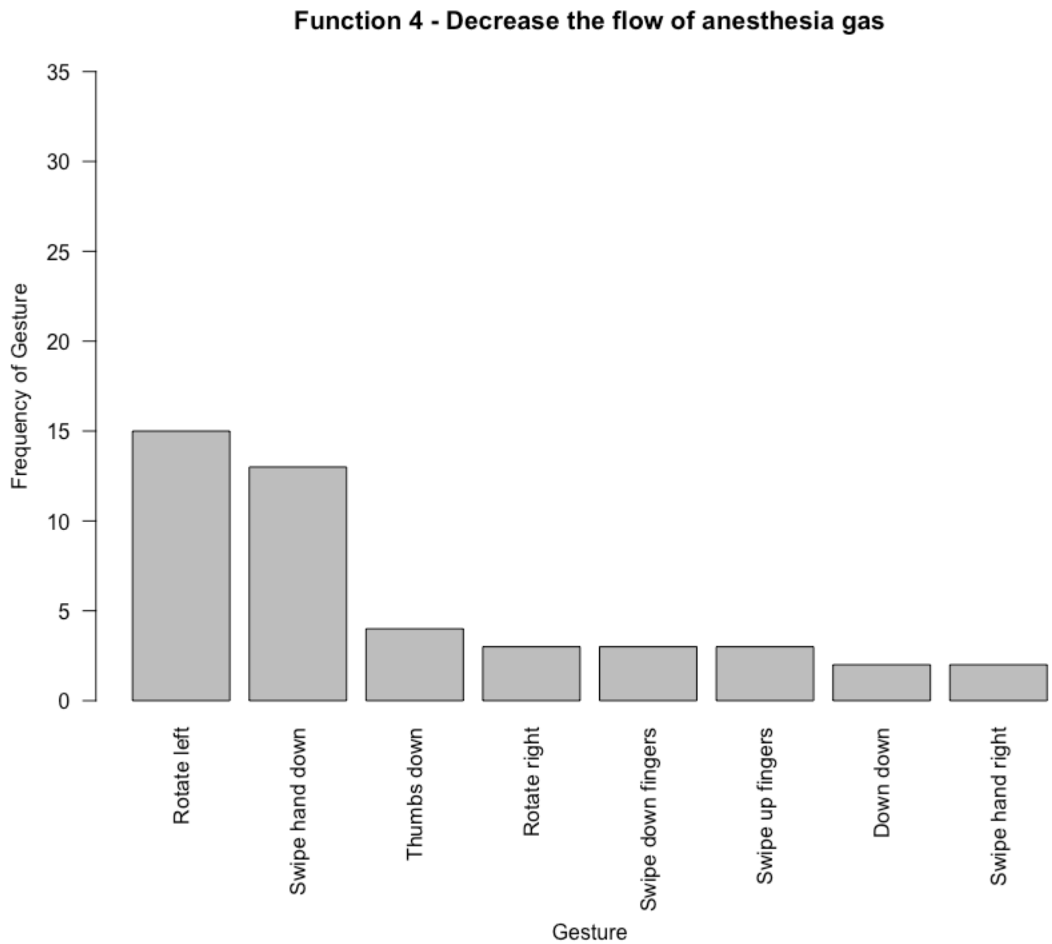
Function 10 - Cancel the request
Frequency of Gesture
0
10
20
30
40
50
60
Swipe left hand
Thumbs down
Swipe right hand
Fist
Push hand
Five up
Wave hand
Click
Thumbs up
Push fingers
X
Come
Okay
Reverse full grab
Two up
Gesture

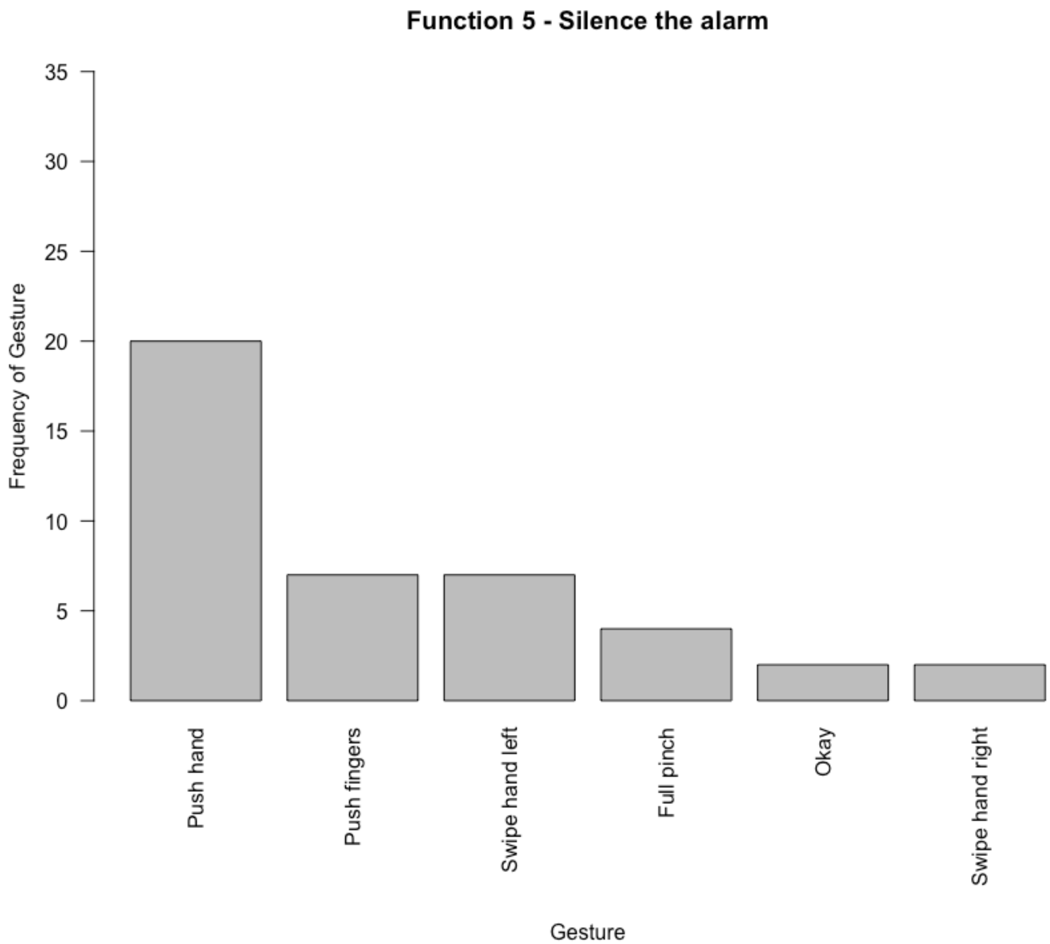
Appendix B - Domain experts bar graphs for intuitive gestures in anesthesia

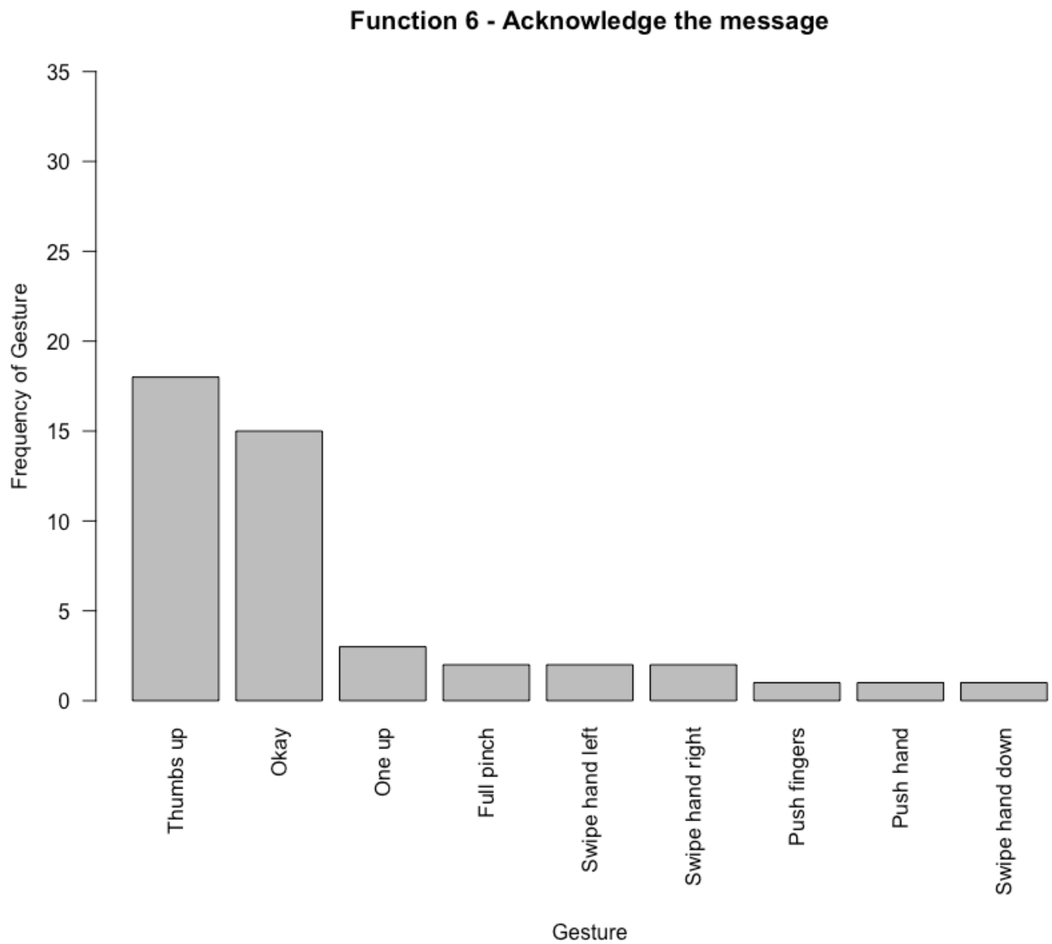


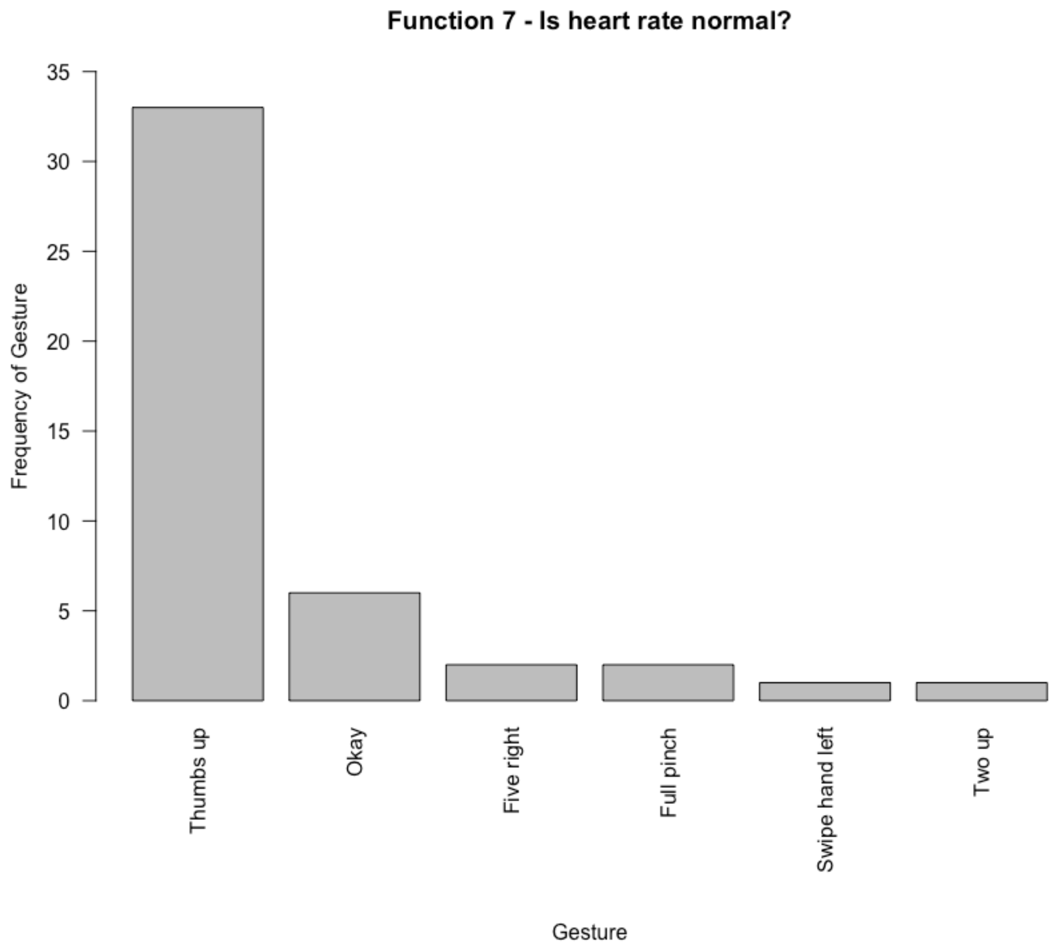

Function 2 - Stop the flow of anesthesia gas
35
30
25
20
15
10
5
0
Frequency of Gesture
Rotate left
Push hand
Fist
Five up
Rotate right
Swipe hand right
Swipe up fingers
Swipe hand down
Swipe hand left
X
Gesture

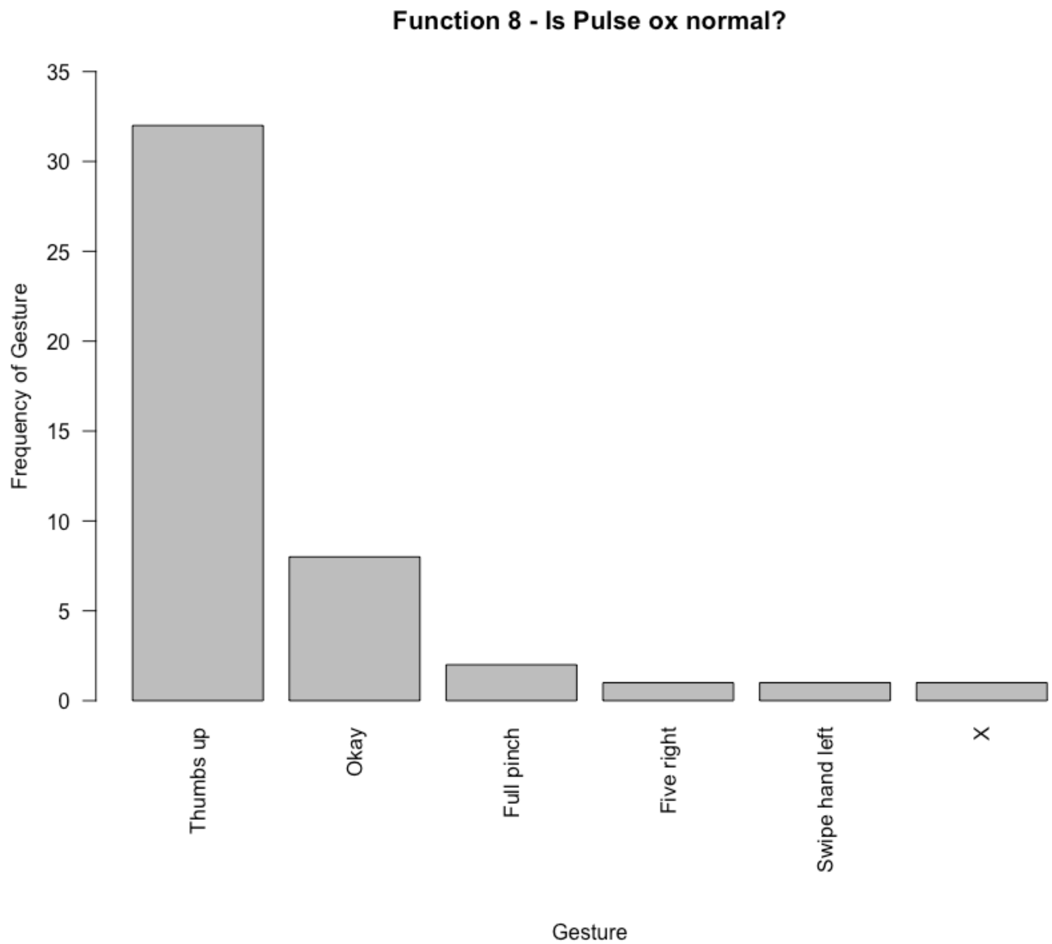

Function 3 - Increase the flow of anesthesia gas
Frequency of Gesture
0
5
10
15
20
25
30
35
Rotate right
Swipe hand up
Swipe up fingers
Rotate left
Swipe hand right
Thumbs up
Up up
Five up
One up
Gesture

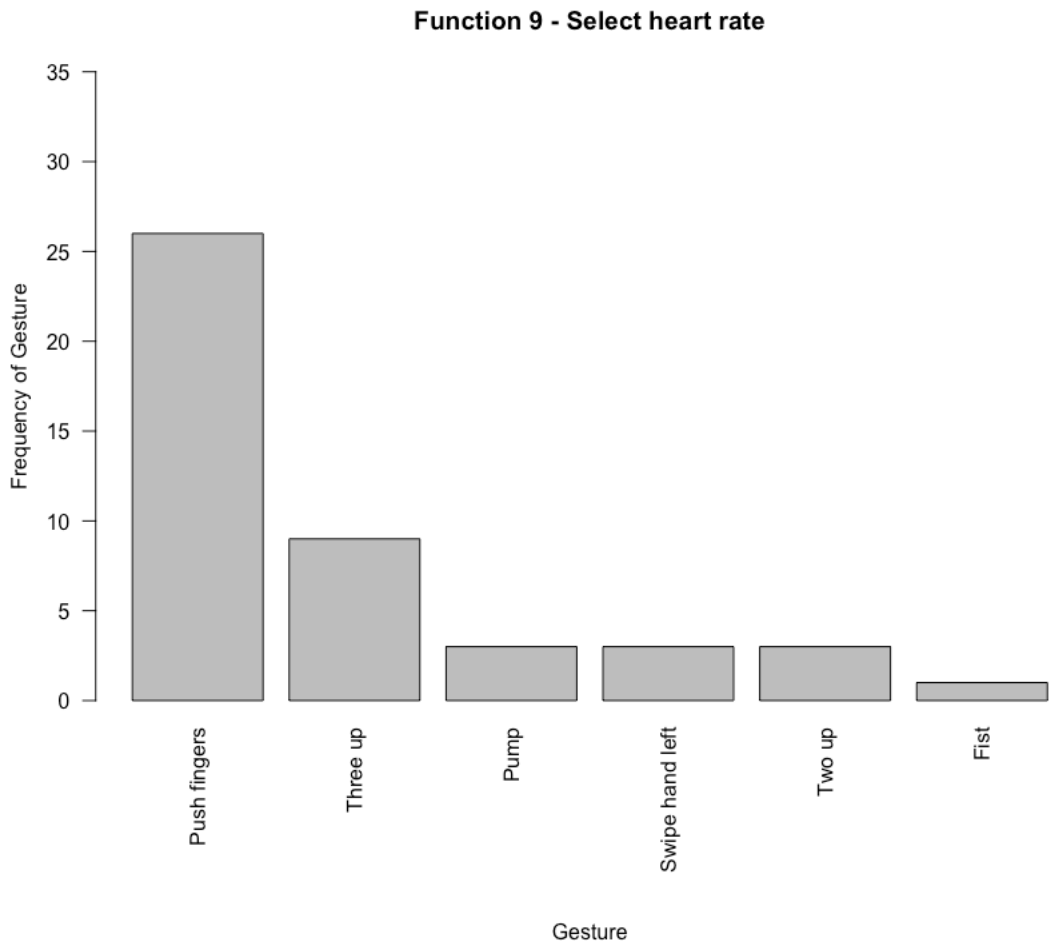

Function 4 - Decrease the flow of anesthesia gas
35
30
25
20
15
10
5
0
Frequency of Gesture
Rotate left
Swipe hand down
Thumbs down
Rotate right
Swipe down fingers
Swipe up fingers
Down down
Swipe hand right
Gesture

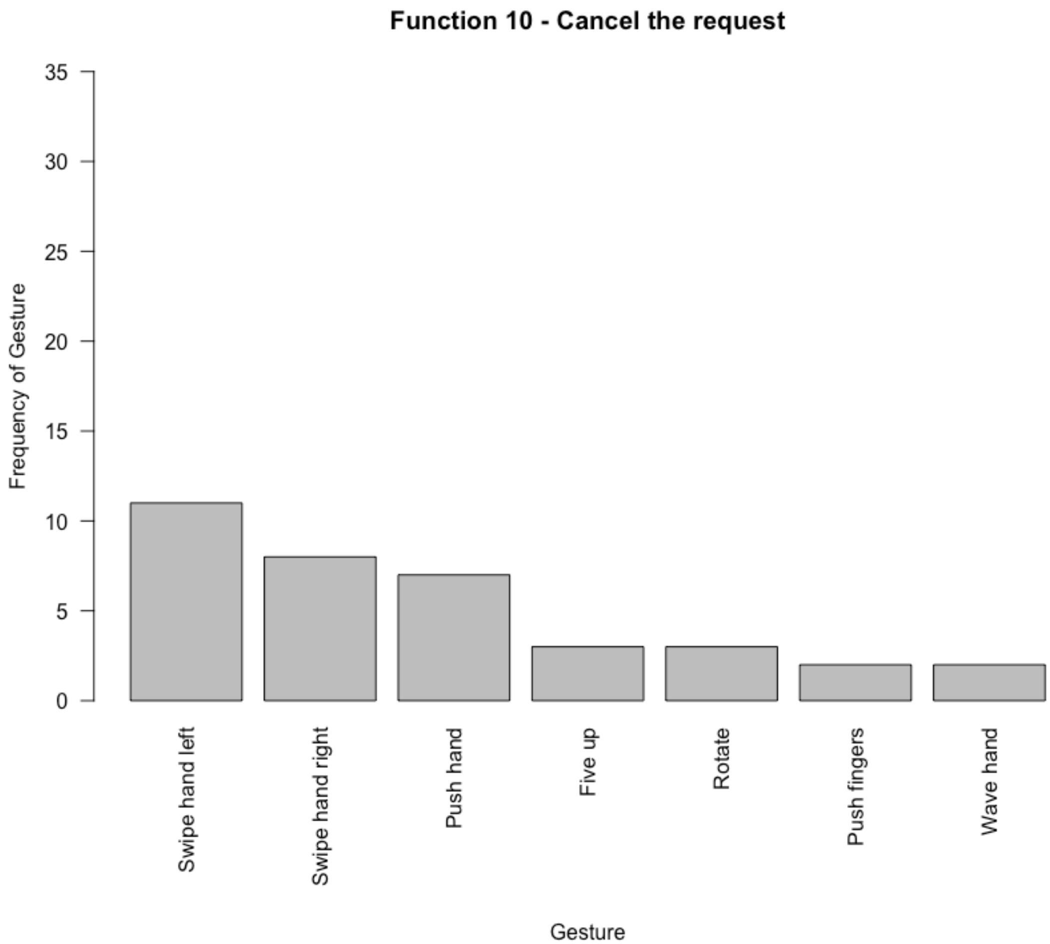

Function 5 - Silence the alarm
Frequency of Gesture
0
5
10
15
20
25
30
35
Push hand
Push fingers
Swipe hand left
Full pinch
Okay
Swipe hand right
Gesture


Function 6 - Acknowledge the message
35
30
25
20
15
10
5
0
Frequency of Gesture
Thumbs up
Okay
One up
Full pinch
Swipe hand left
Swipe hand right
Push fingers
Push hand
Swipe hand down
Gesture


Function 7 - Is heart rate normal?
Frequency of Gesture
35
30
25
20
15
10
5
0
Thumbs up
Okay
Five right
Full pinch
Swipe hand left
Two up
Gesture

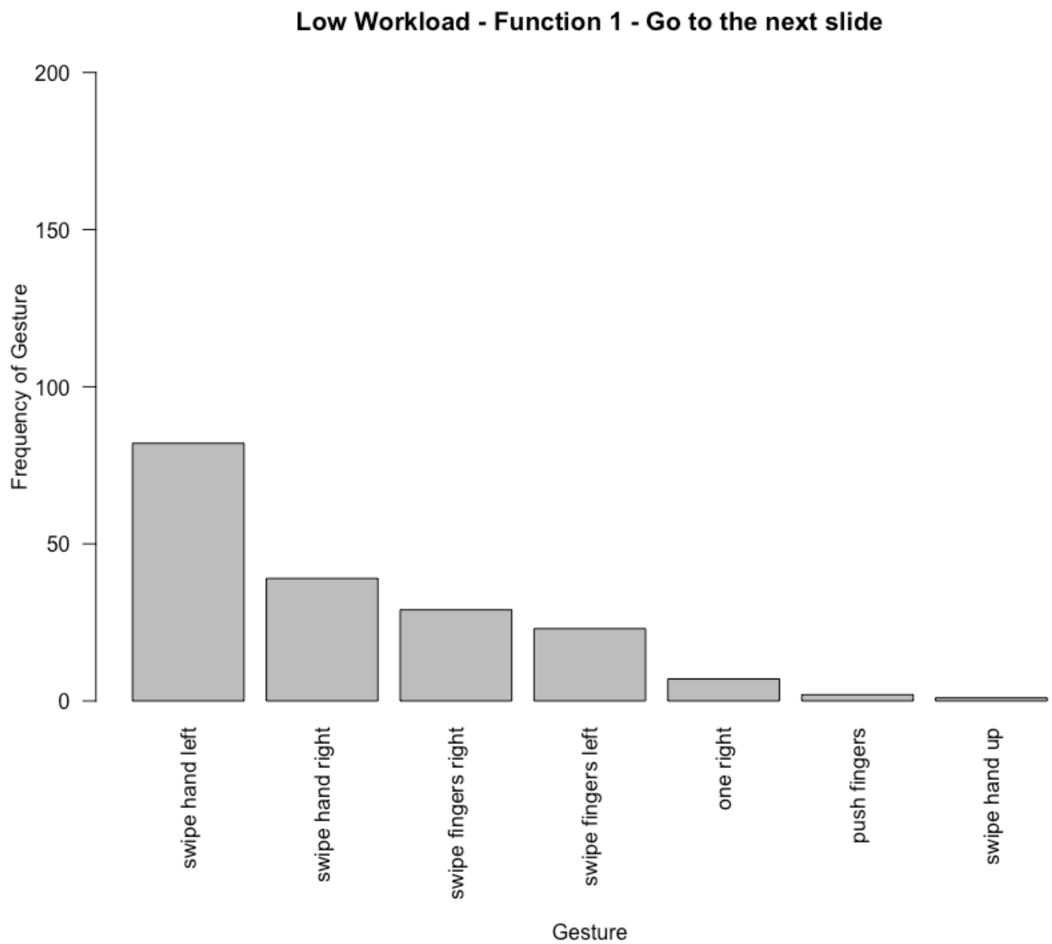

Function 8 - Is Pulse ox normal?
Frequency of Gesture
35
30
25
20
15
10
5
0
Thumbs up
Okay
Full pinch
Five right
Swipe hand left
X
Gesture

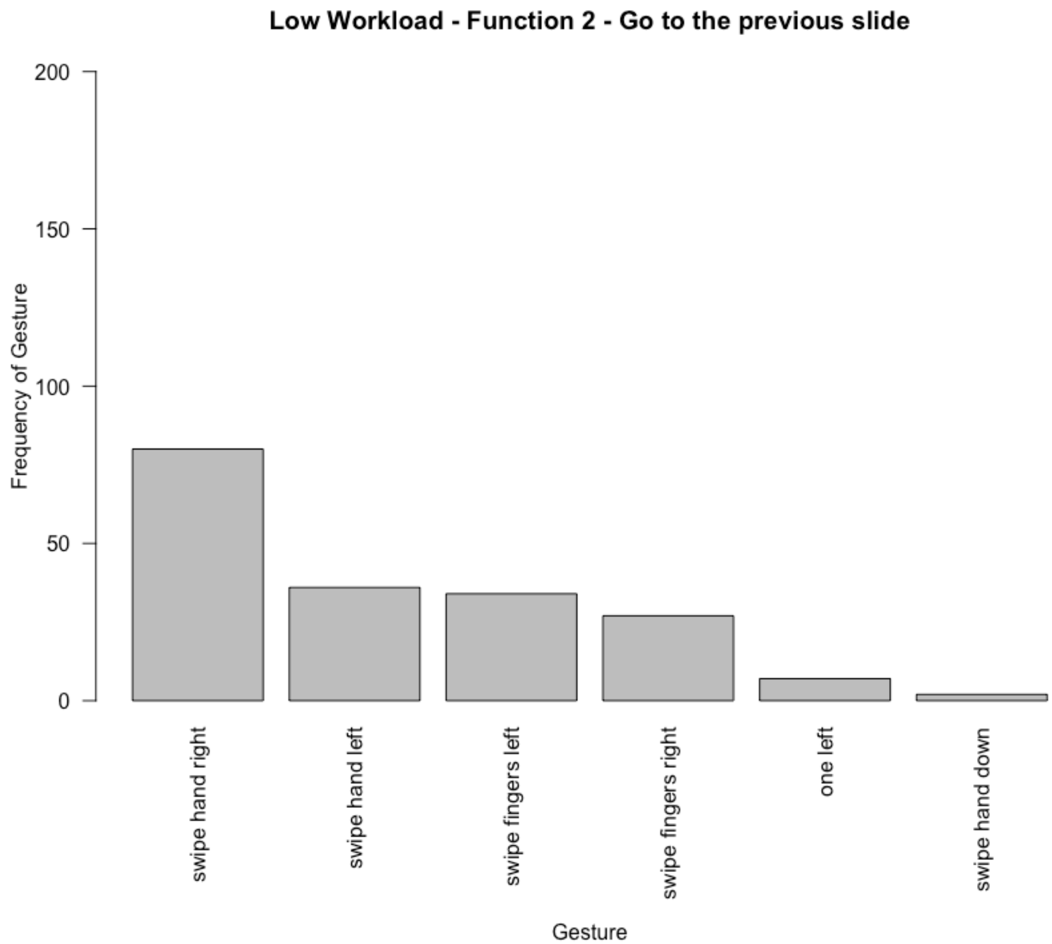

Function 9 - Select heart rate
Frequency of Gesture
35
30
25
20
15
10
5
0
Push fingers
Three up
Pump
Swipe hand left
Two up
Fist
Gesture



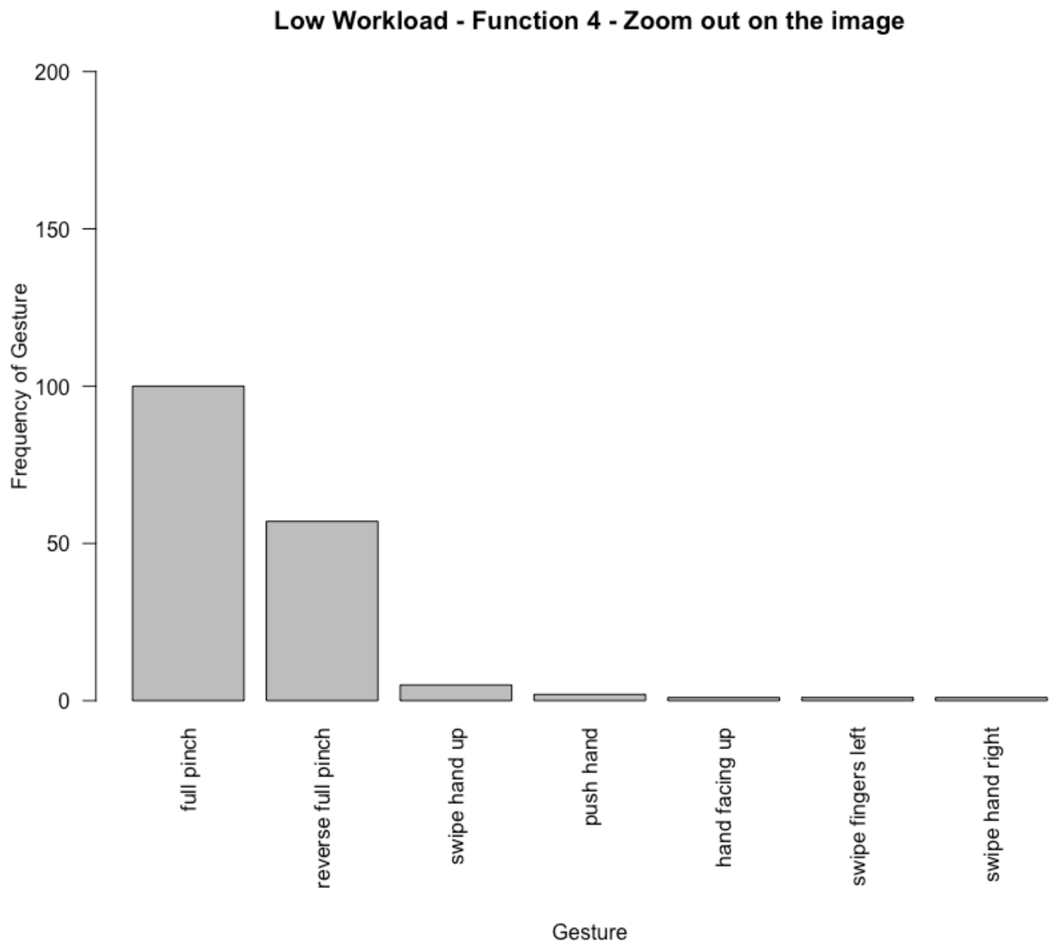
Function 10 - Cancel the request
Frequency of Gesture
35
30
25
20
15
10
5
0
Swipe hand left
Swipe hand right
Push hand
Five up
Rotate
Push fingers
Wave hand
Gesture

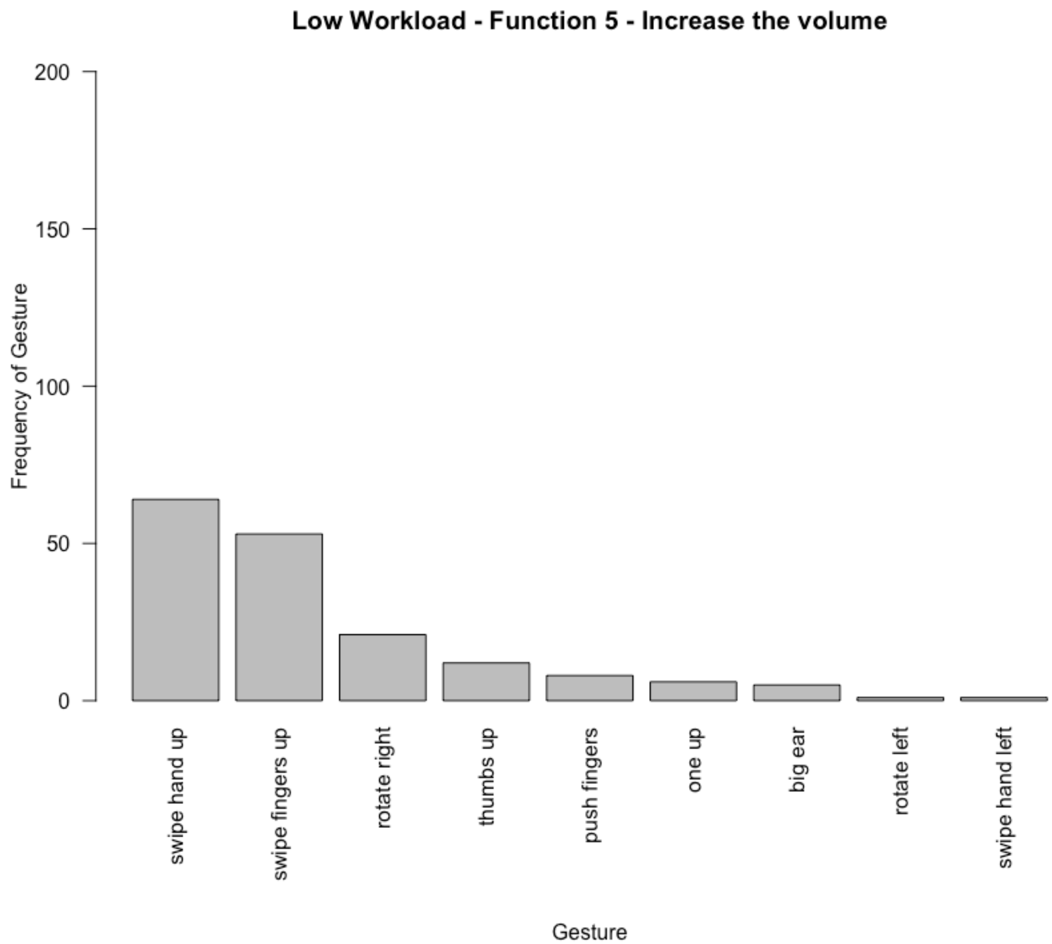
Appendix C – Low workload bar graphs for intuitive gestures for controlling a presentation

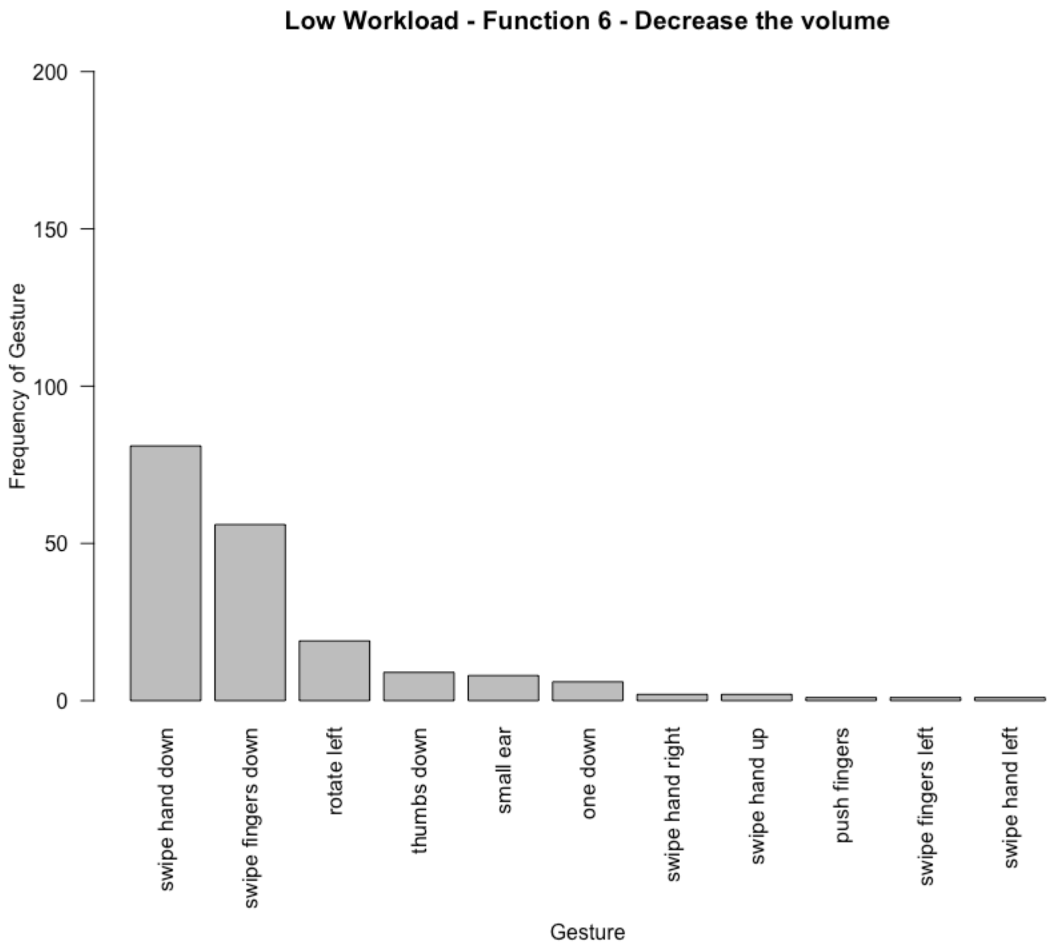


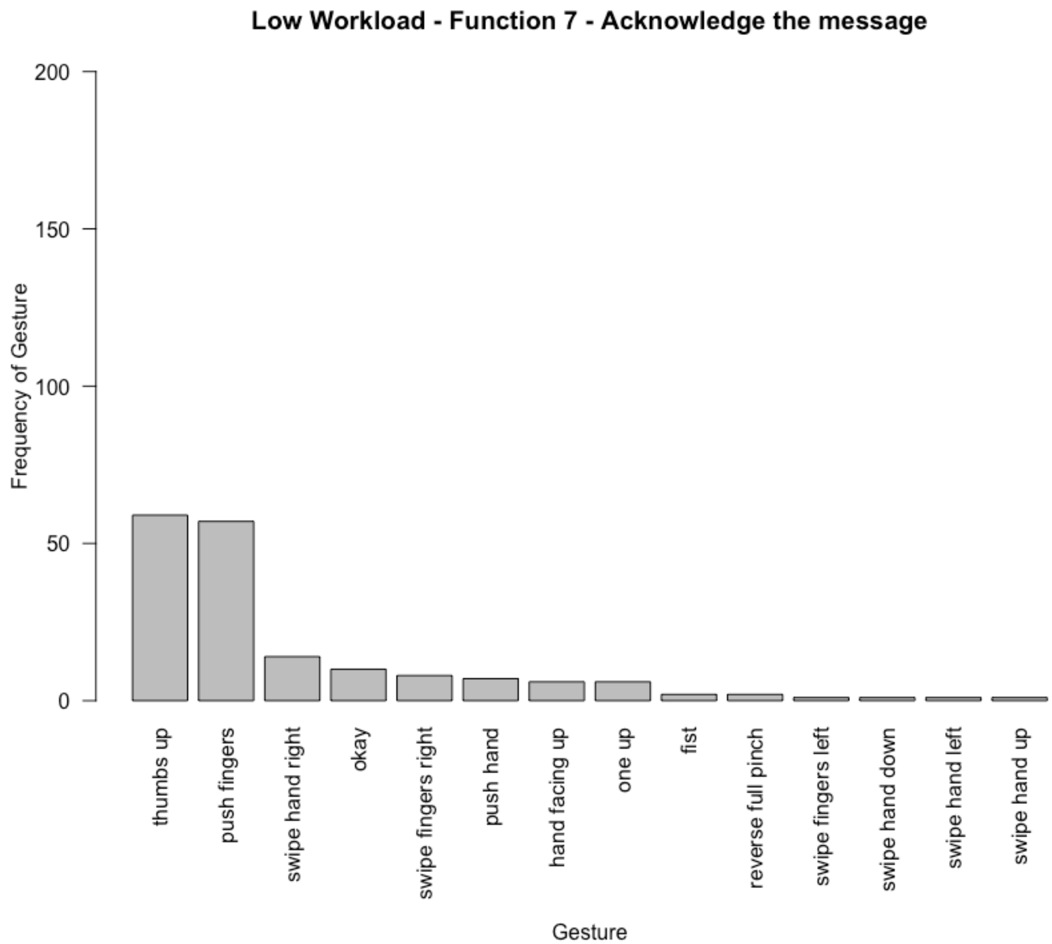

Low Workload - Function 2 - Go to the previous slide
Frequency of Gesture
200
150
100
50
0
swipe hand right
swipe hand left
swipe fingers left
swipe fingers right
one left
swipe hand down
Gesture

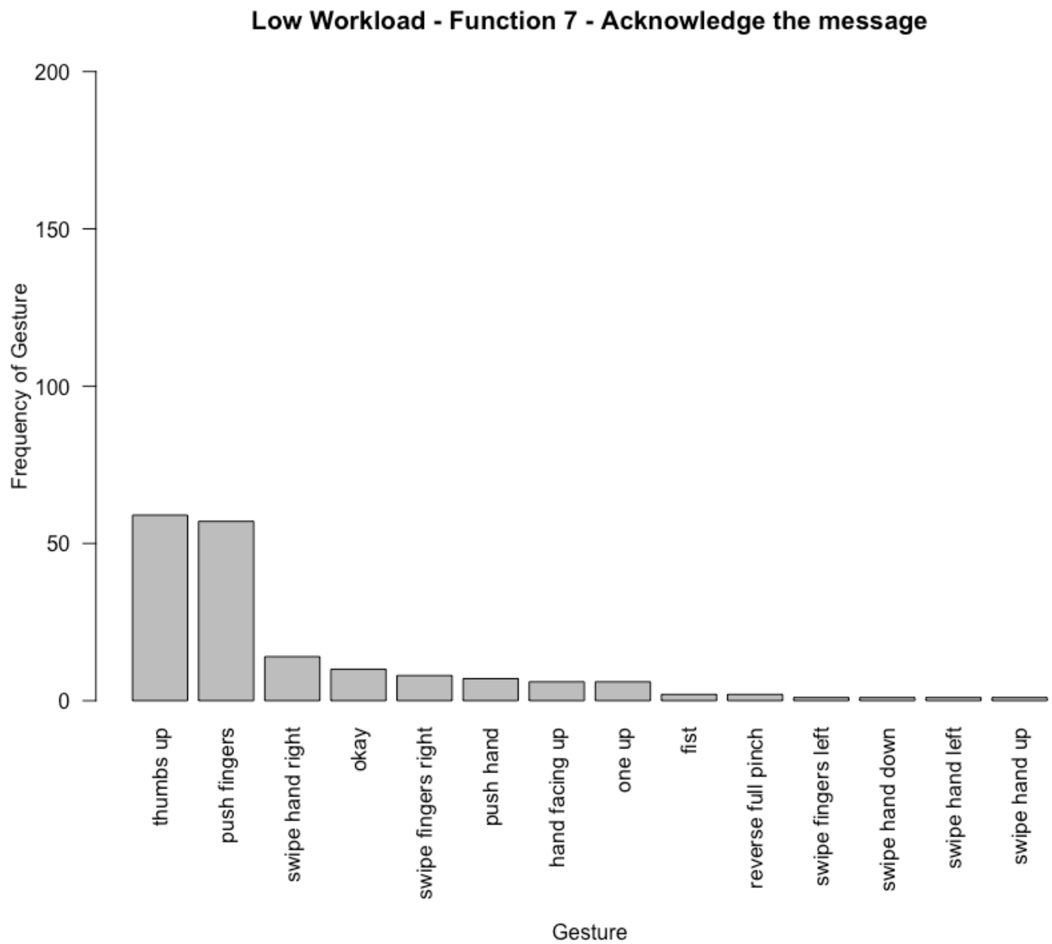

Low Workload - Function 3 - Zoom in on the image
Frequency of Gesture
200
150
100
50
0
reverse full pinch
full pinch
swipe fingers left
push fingers
swipe hand right
swipe hand up
Gesture


Low Workload - Function 4 - Zoom out on the image
Frequency of Gesture
200
150
100
50
0
full pinch
reverse full pinch
swipe hand up
push hand
hand facing up
swipe fingers left
swipe hand right
Gesture

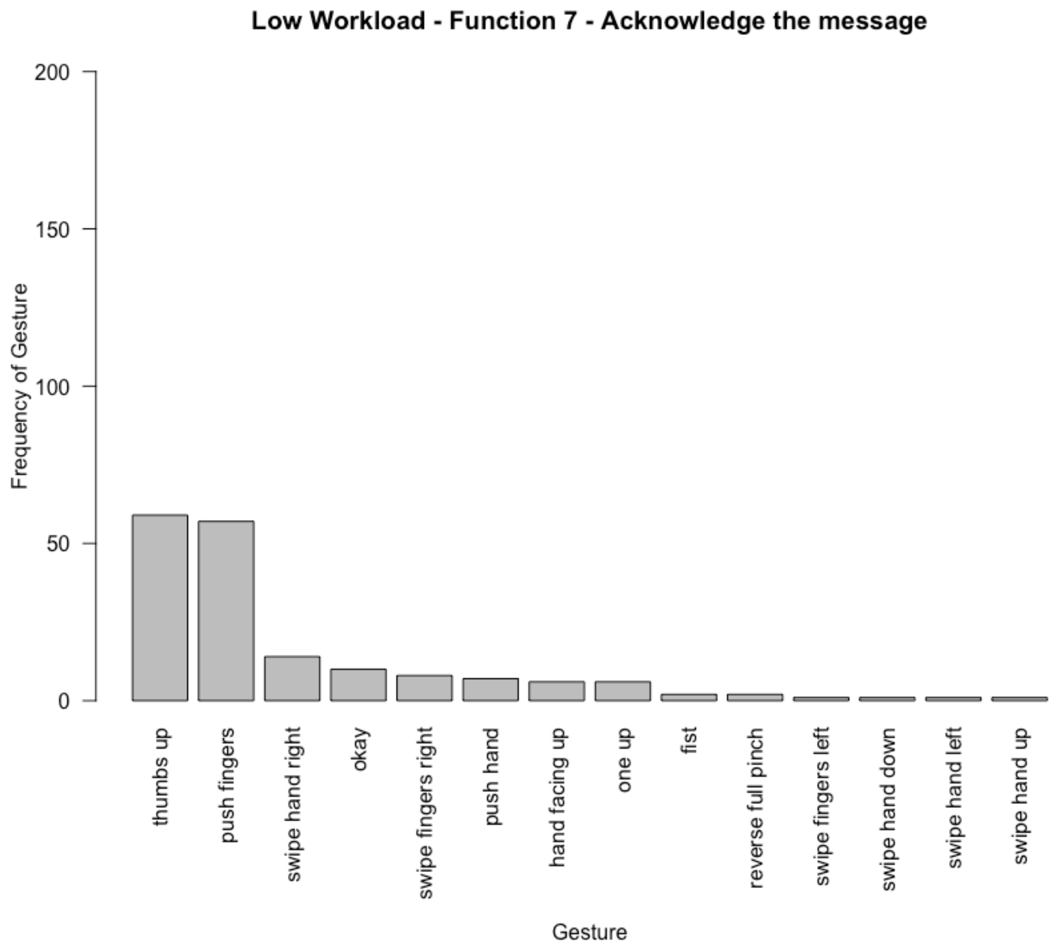

Low Workload - Function 5 - Increase the volume
200
150
100
50
0
Frequency of Gesture
swipe hand up
swipe fingers up
rotate right
thumbs up
push fingers
one up
big ear
rotate left
swipe hand left
Gesture


Low Workload - Function 6 - Decrease the volume
200
150
100
50
0
Frequency of Gesture
swipe hand down
swipe fingers down
rotate left
thumbs down
small ear
one down
swipe hand right
swipe hand up
push fingers
swipe fingers left
swipe hand left
Gesture

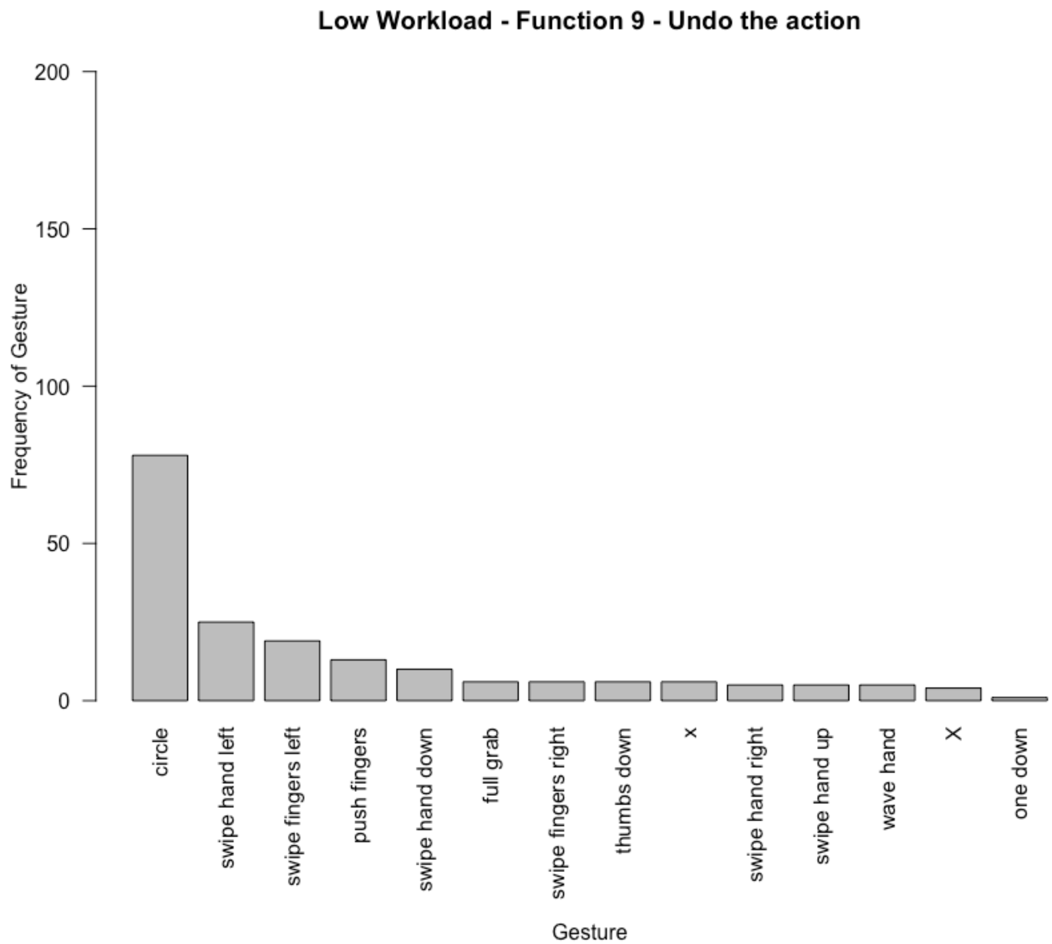

Low Workload - Function 7 - Acknowledge the message
Frequency of Gesture
200
150
100
50
0
thumbs up
push fingers
swipe hand right
okay
swipe fingers right
push hand
hand facing up
one up
fist
reverse full pinch
swipe fingers left
swipe hand down
swipe hand left
swipe hand up
Gesture



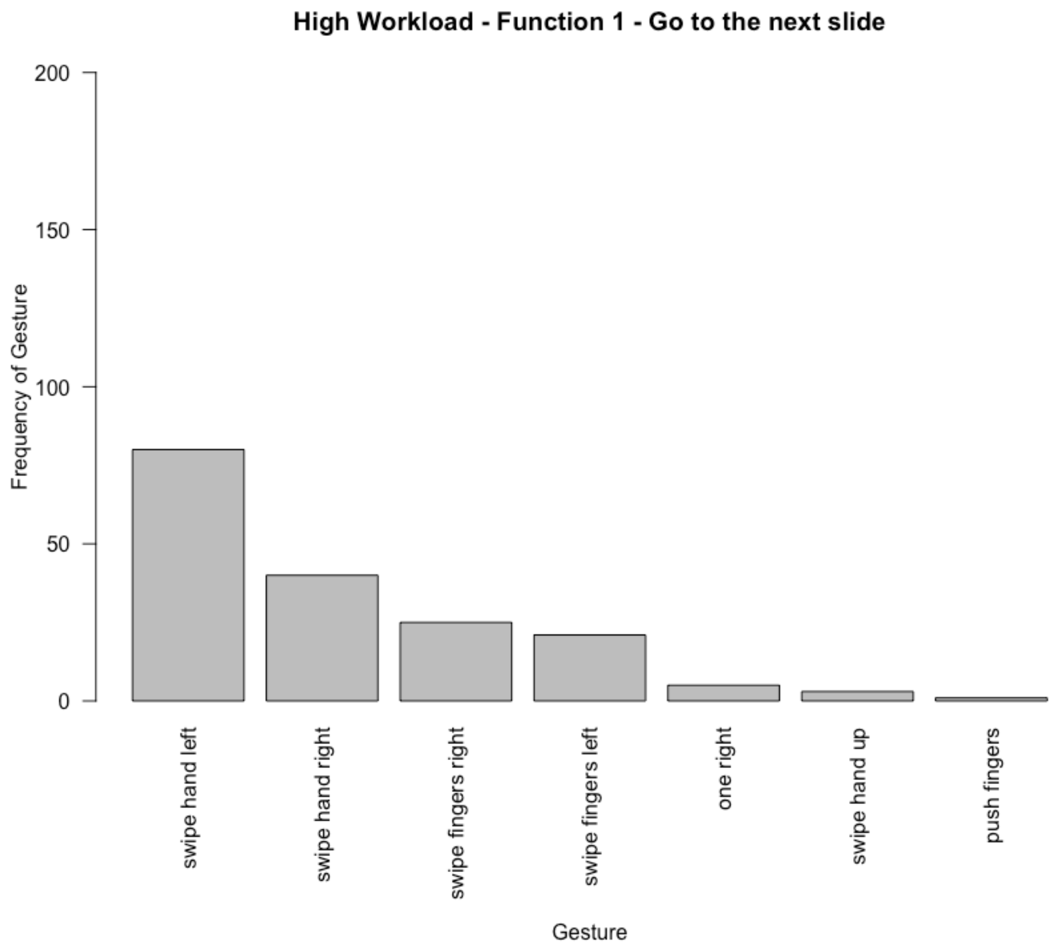
Low Workload - Function 7 - Acknowledge the message
Frequency of Gesture
200
150
100
50
0
thumbs up
push fingers
swipe hand right
okay
swipe fingers right
push hand
hand facing up
one up
fist
reverse full pinch
swipe fingers left
swipe hand down
swipe hand left
swipe hand up
Gesture

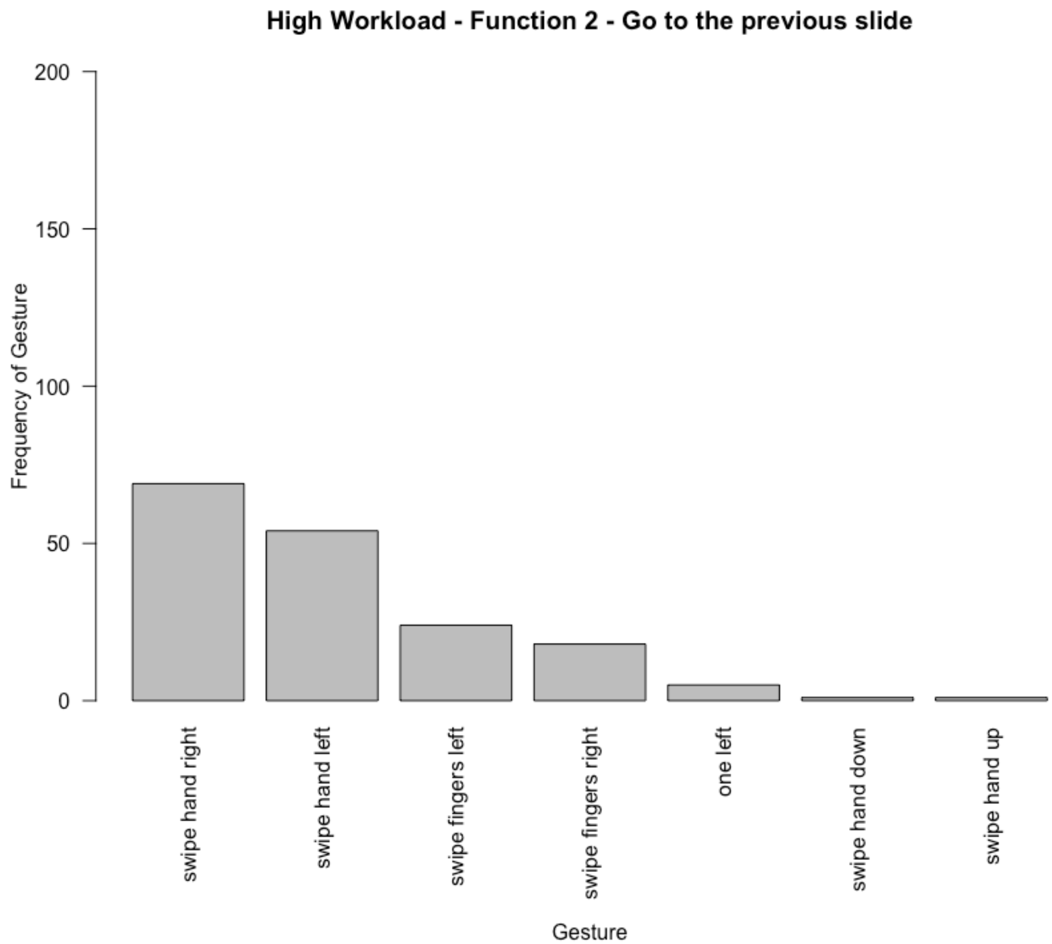


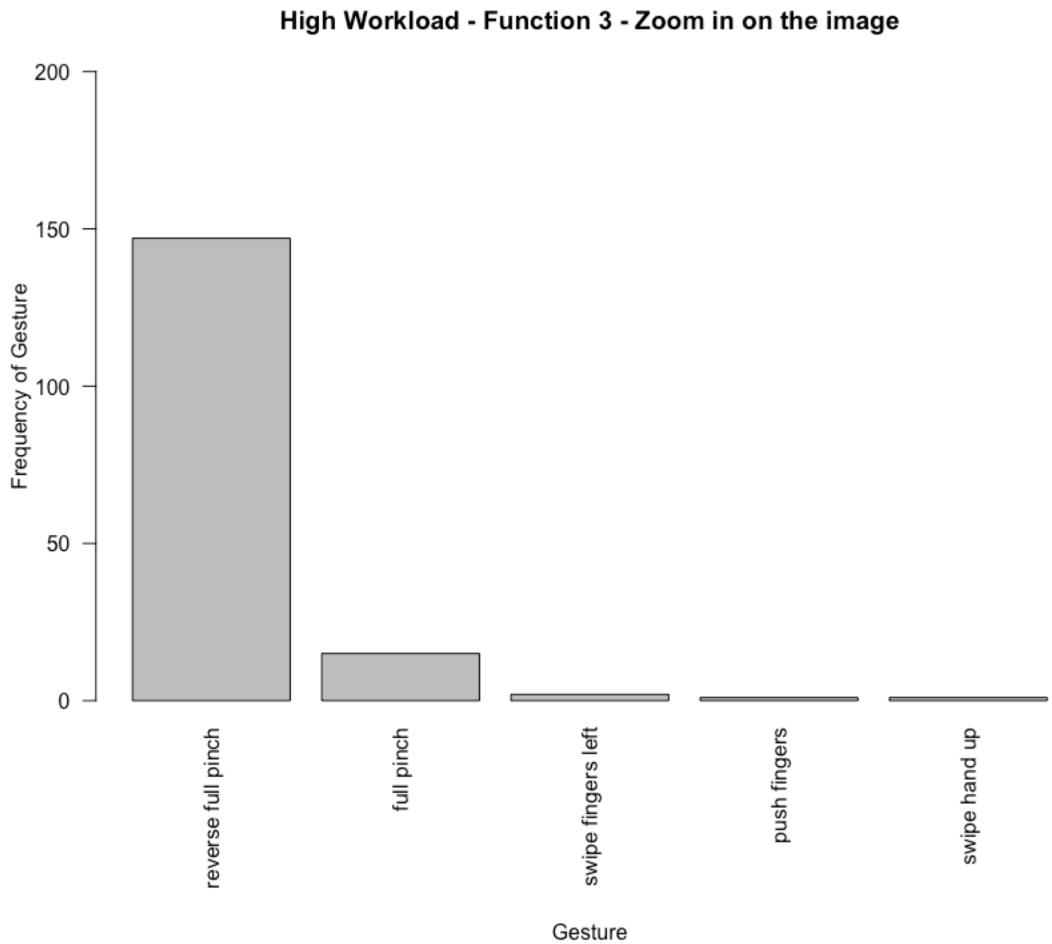
Low Workload - Function 7 - Acknowledge the message
Frequency of Gesture
200
150
100
50
0
thumbs up
push fingers
swipe hand right
okay
swipe fingers right
push hand
hand facing up
one up
fist
reverse full pinch
swipe fingers left
swipe hand down
swipe hand left
swipe hand up
Gesture

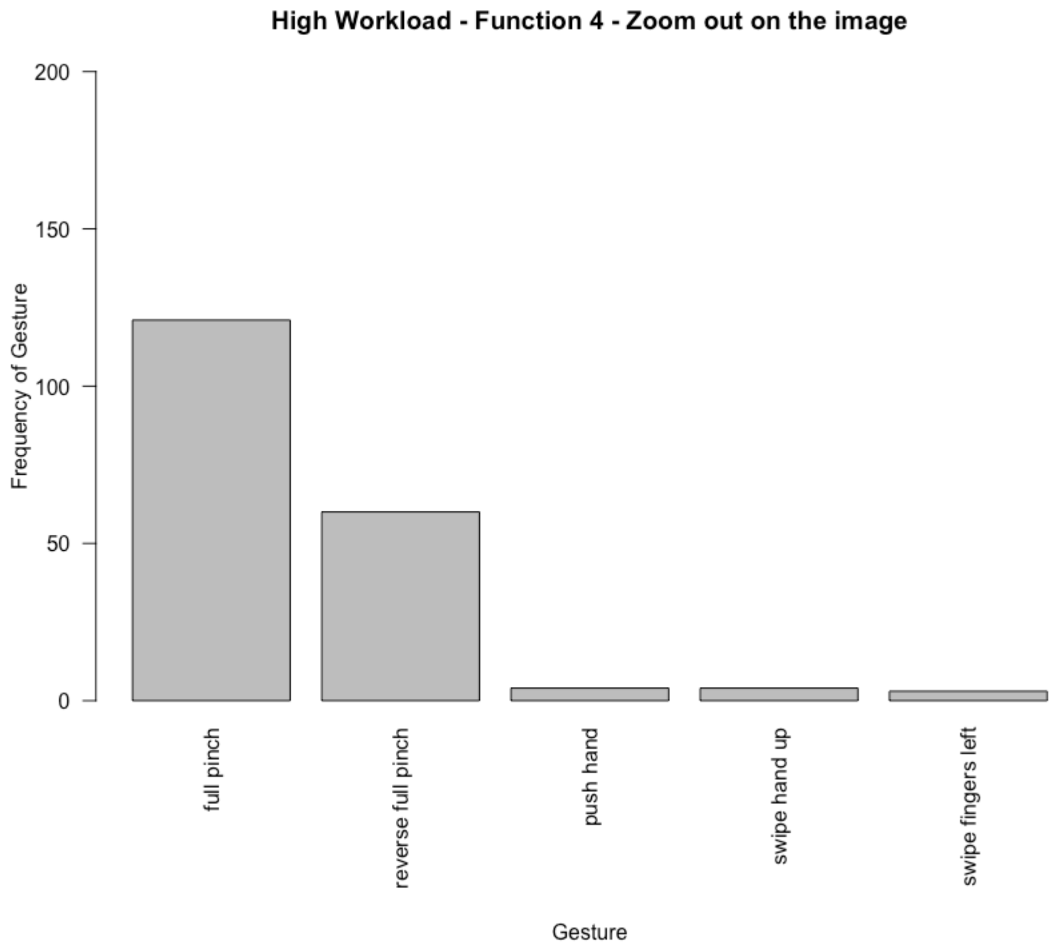

Low Workload - Function 9 - Undo the action
Frequency of Gesture
0
50
100
150
200
circle
swipe hand left
swipe fingers left
push fingers
swipe hand down
full grab
swipe fingers right
thumbs down
x
swipe hand right
swipe hand up
wave hand
X
one down
Gesture

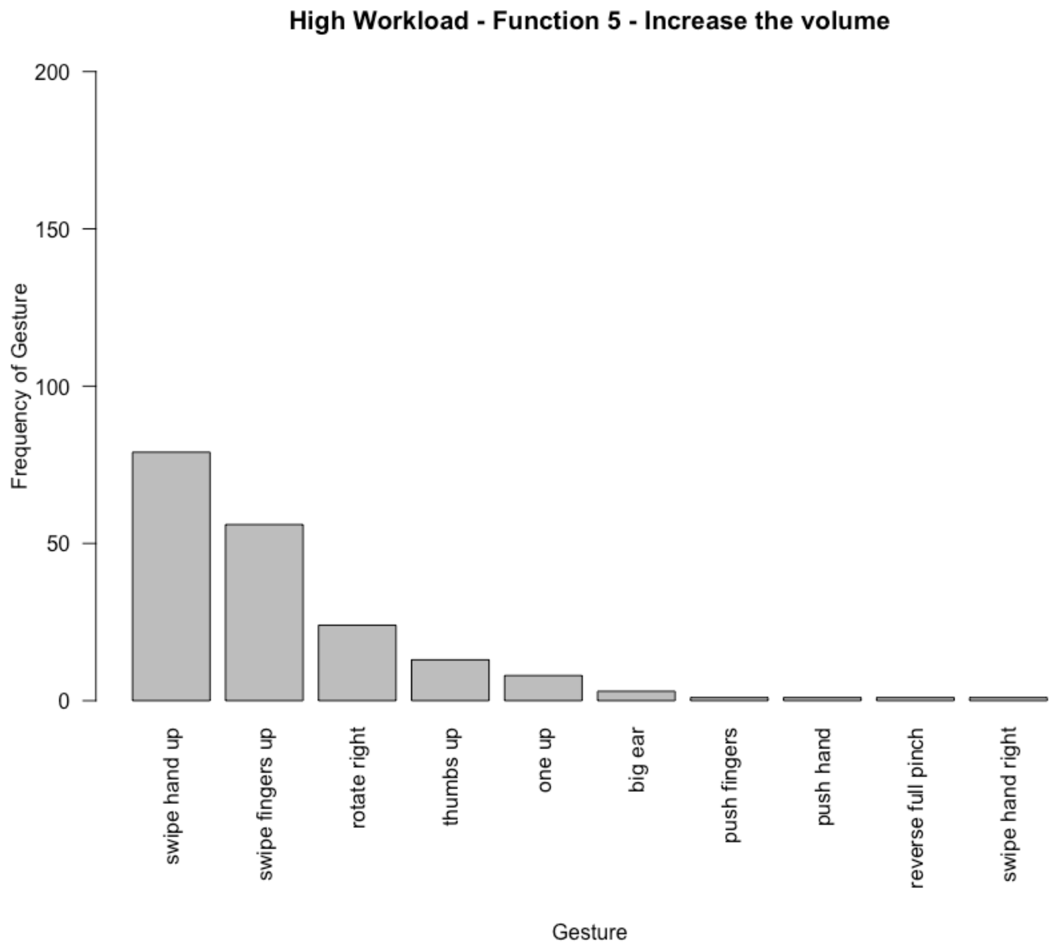
Appendix D -High workload bar graphs for intuitive gestures for controlling a presentation

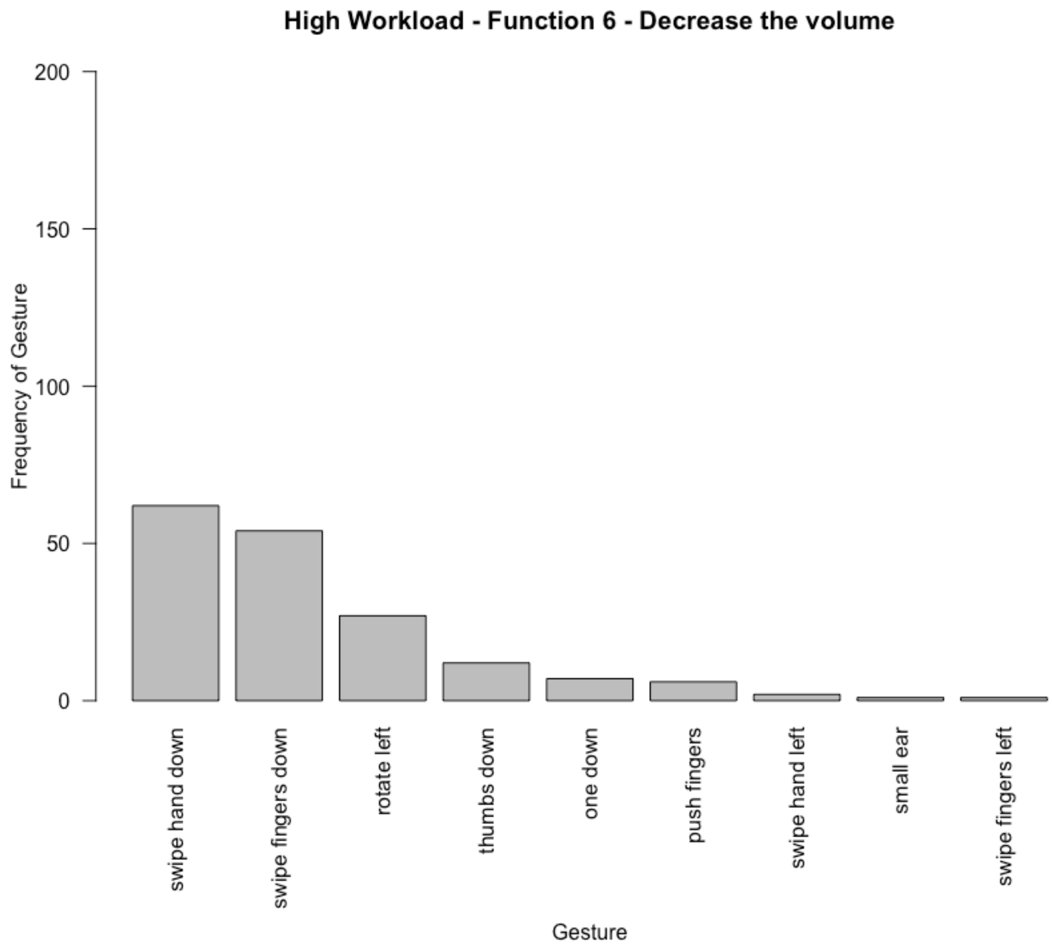


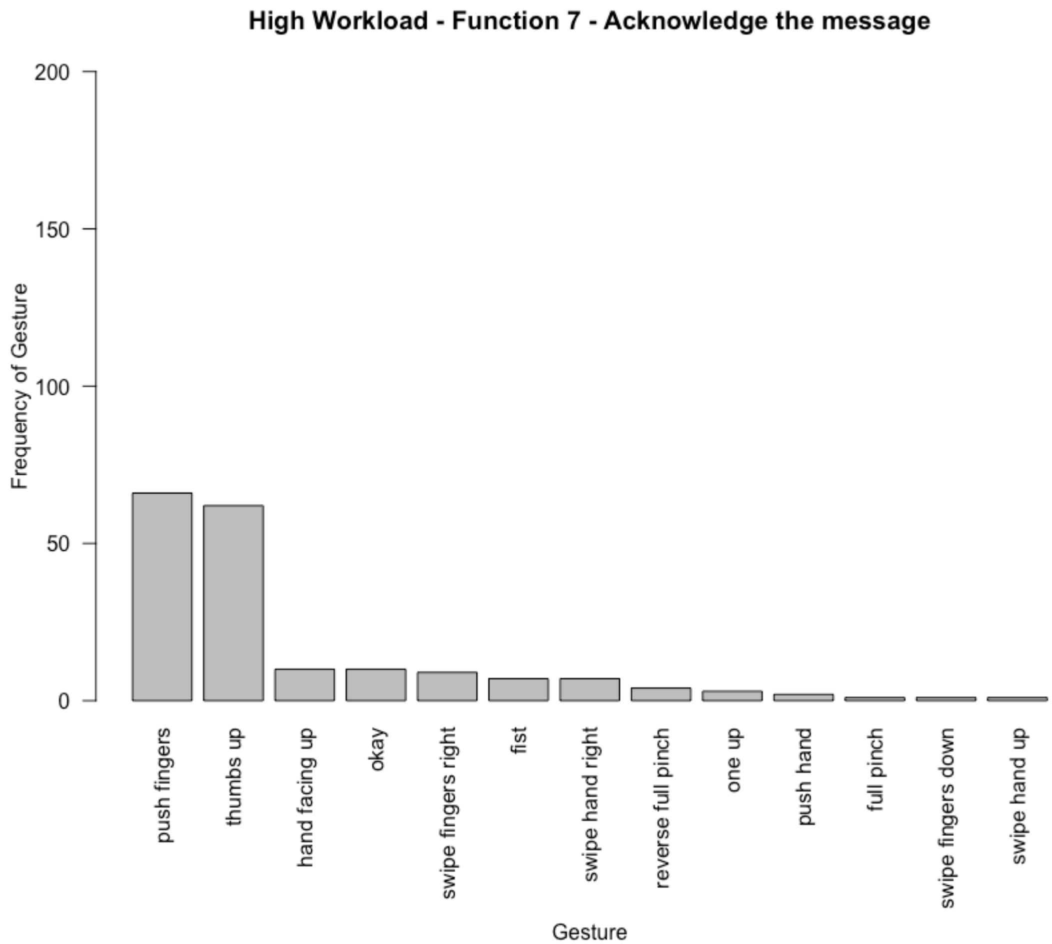

High Workload - Function 2 - Go to the previous slide
Frequency of Gesture
200
150
100
50
0
swipe hand right
swipe hand left
swipe fingers left
swipe fingers right
one left
swipe hand down
swipe hand up
Gesture

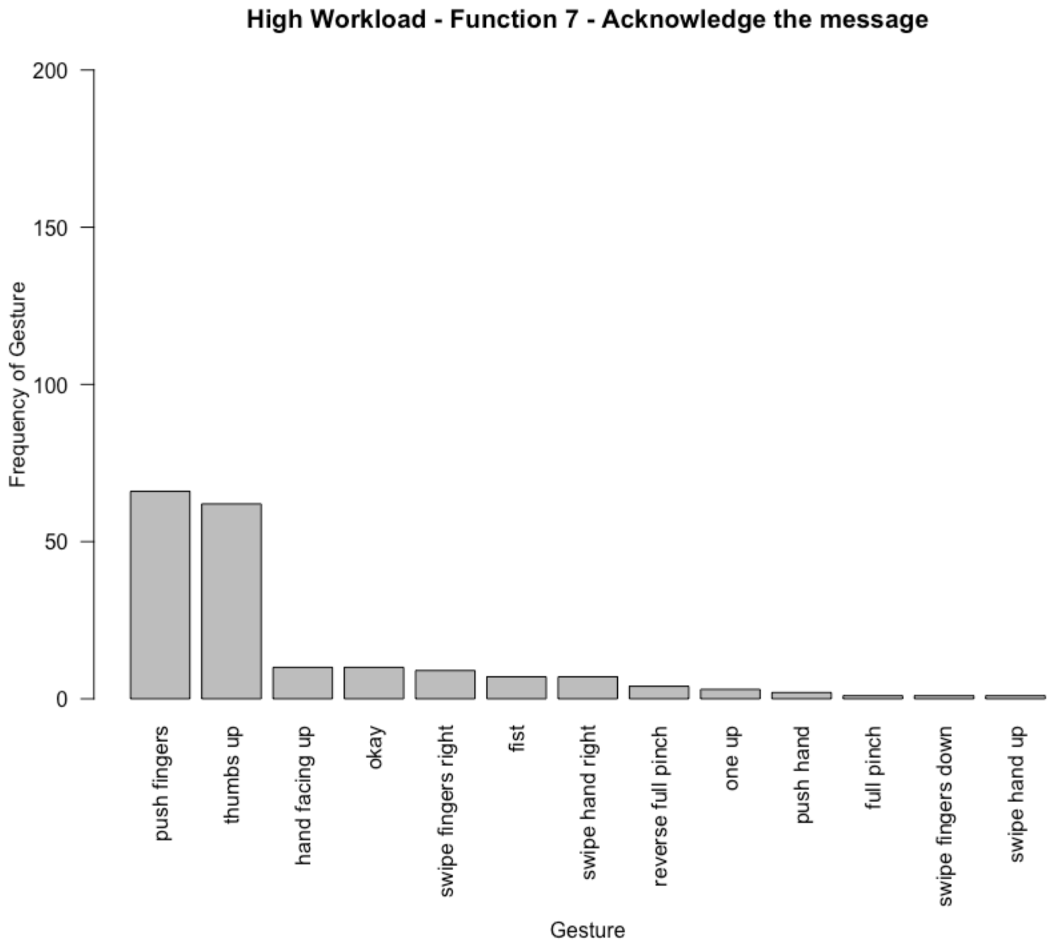

High Workload - Function 3 - Zoom in on the image
200
150
100
50
0
Frequency of Gesture
reverse full pinch
full pinch
swipe fingers left
push fingers
swipe hand up
Gesture

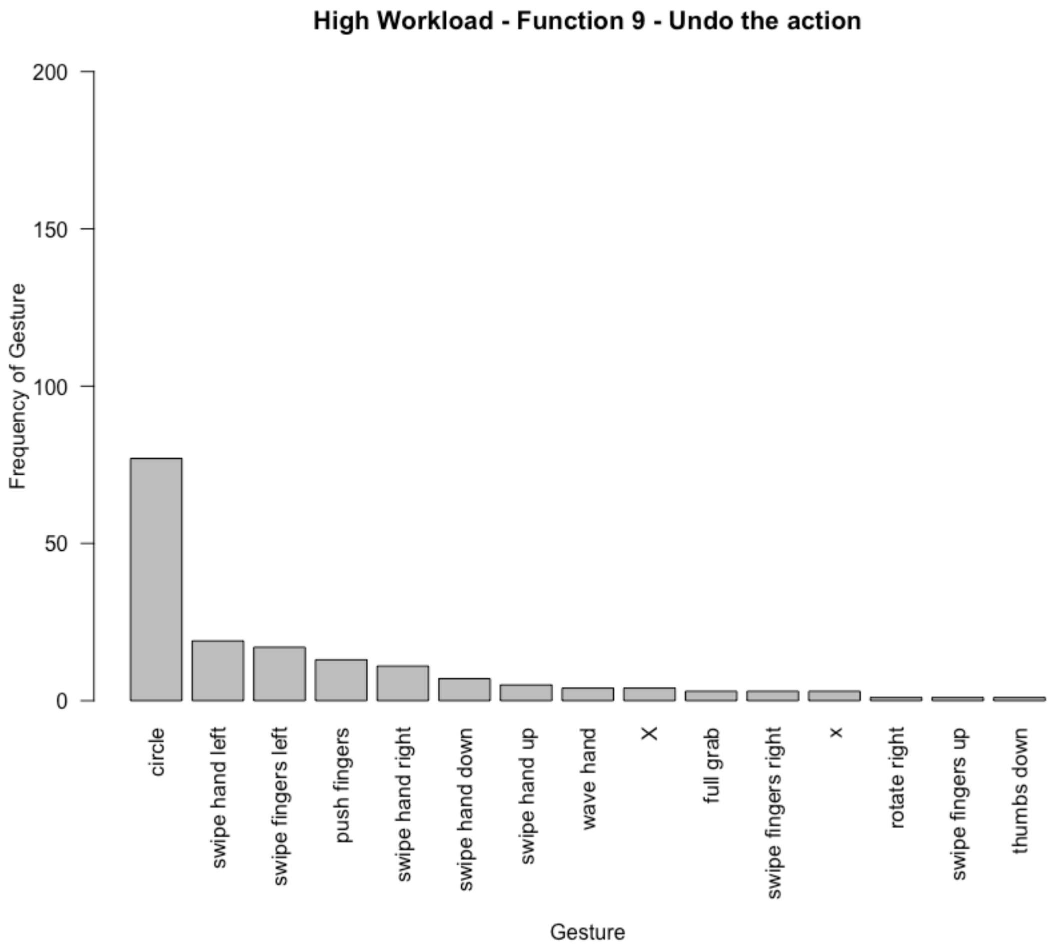

High Workload - Function 4 - Zoom out on the image
Frequency of Gesture
200
150
100
50
0
full pinch
reverse full pinch
push hand
swipe hand up
swipe fingers left
Gesture


High Workload - Function 5 - Increase the volume
Frequency of Gesture
200
150
100
50
0
swipe hand up
swipe fingers up
rotate right
thumbs up
one up
big ear
push fingers
push hand
reverse full pinch
swipe hand right
Gesture

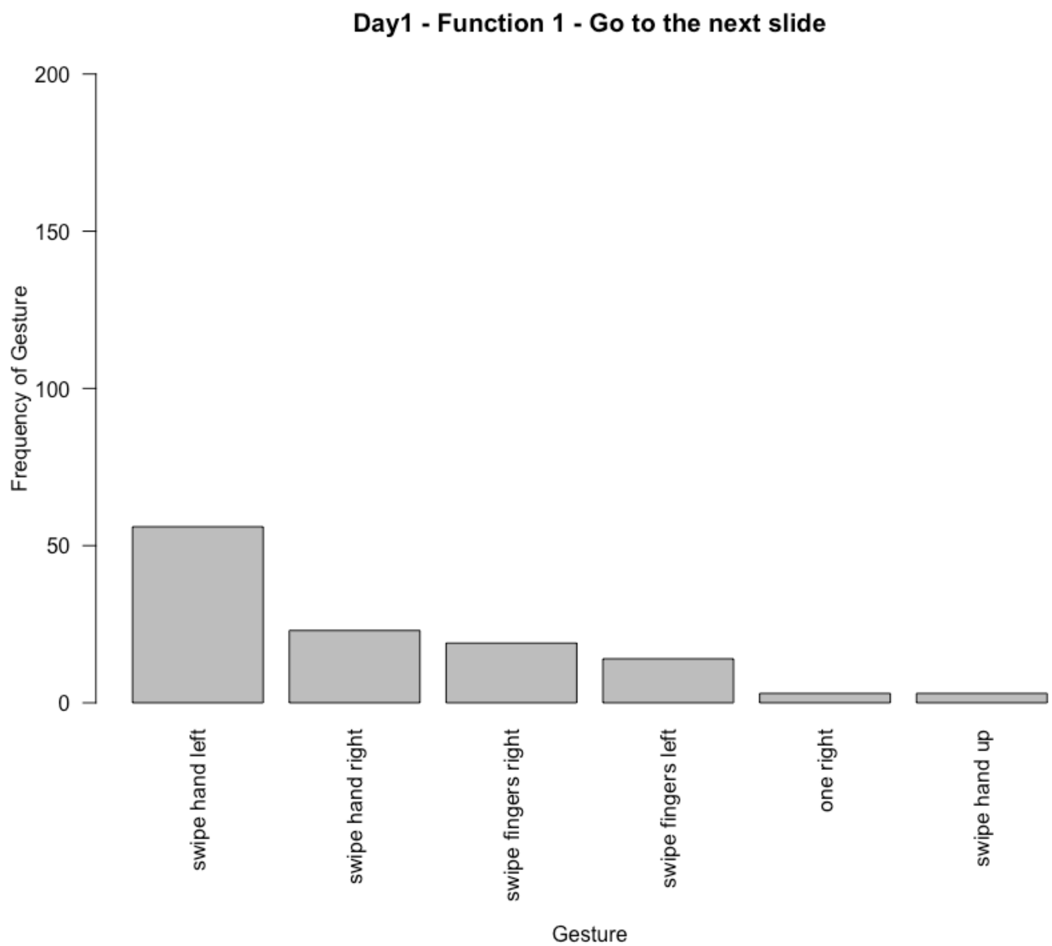

High Workload - Function 6 - Decrease the volume
Frequency of Gesture
200
150
100
50
0
swipe hand down
swipe fingers down
rotate left
thumbs down
one down
push fingers
swipe hand left
small ear
swipe fingers left
Gesture

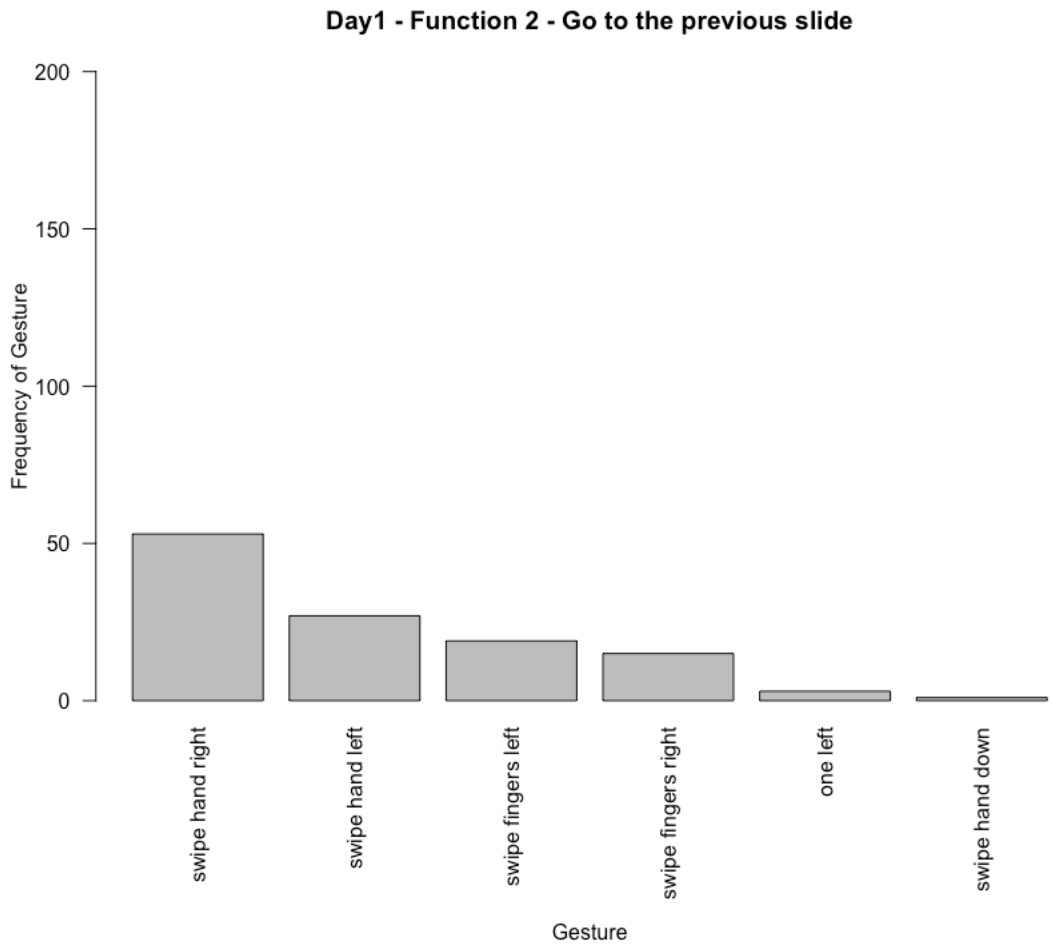

High Workload - Function 7 - Acknowledge the message
200
150
100
50
0
Frequency of Gesture
push fingers
thumbs up
hand facing up
okay
swipe fingers right
fist
swipe hand right
reverse full pinch
one up
push hand
full pinch
swipe fingers down
swipe hand up
Gesture

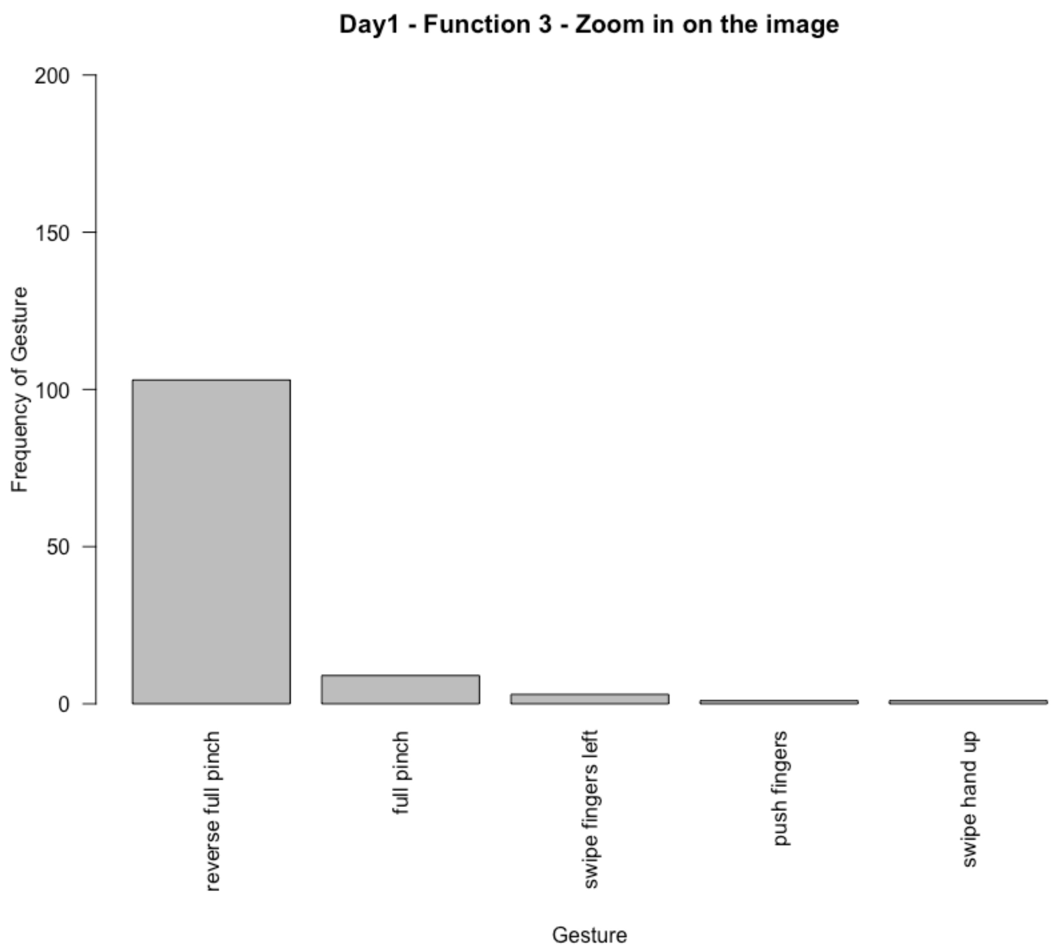

High Workload - Function 7 - Acknowledge the message
Frequency of Gesture
200
150
100
50
0
push fingers
thumbs up
hand facing up
okay
swipe fingers right
fist
swipe hand right
reverse full pinch
one up
push hand
full pinch
swipe fingers down
swipe hand up
Gesture

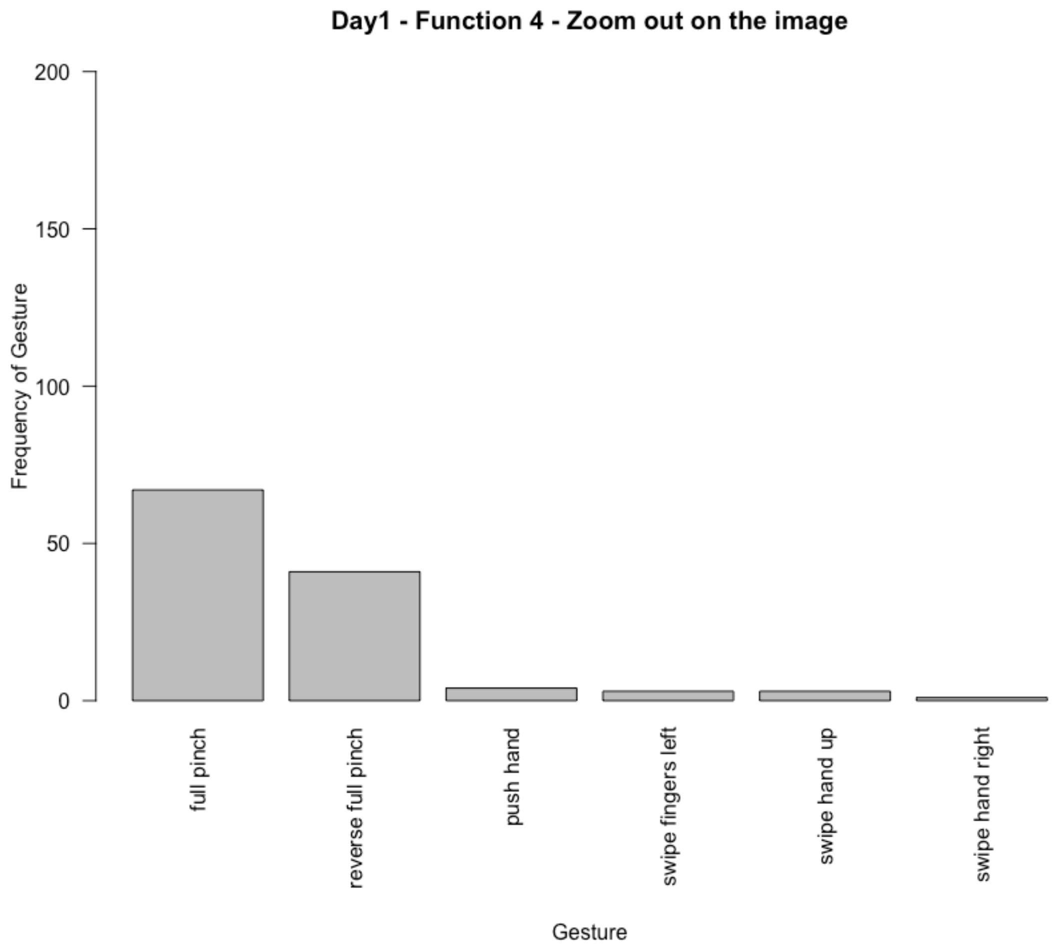

High Workload - Function 9 - Undo the action
Frequency of Gesture
200
150
100
50
0
circle
swipe hand left
swipe fingers left
push fingers
swipe hand right
swipe hand down
swipe hand up
wave hand
X
full grab
swipe fingers right
x
rotate right
swipe fingers up
thumbs down
Gesture

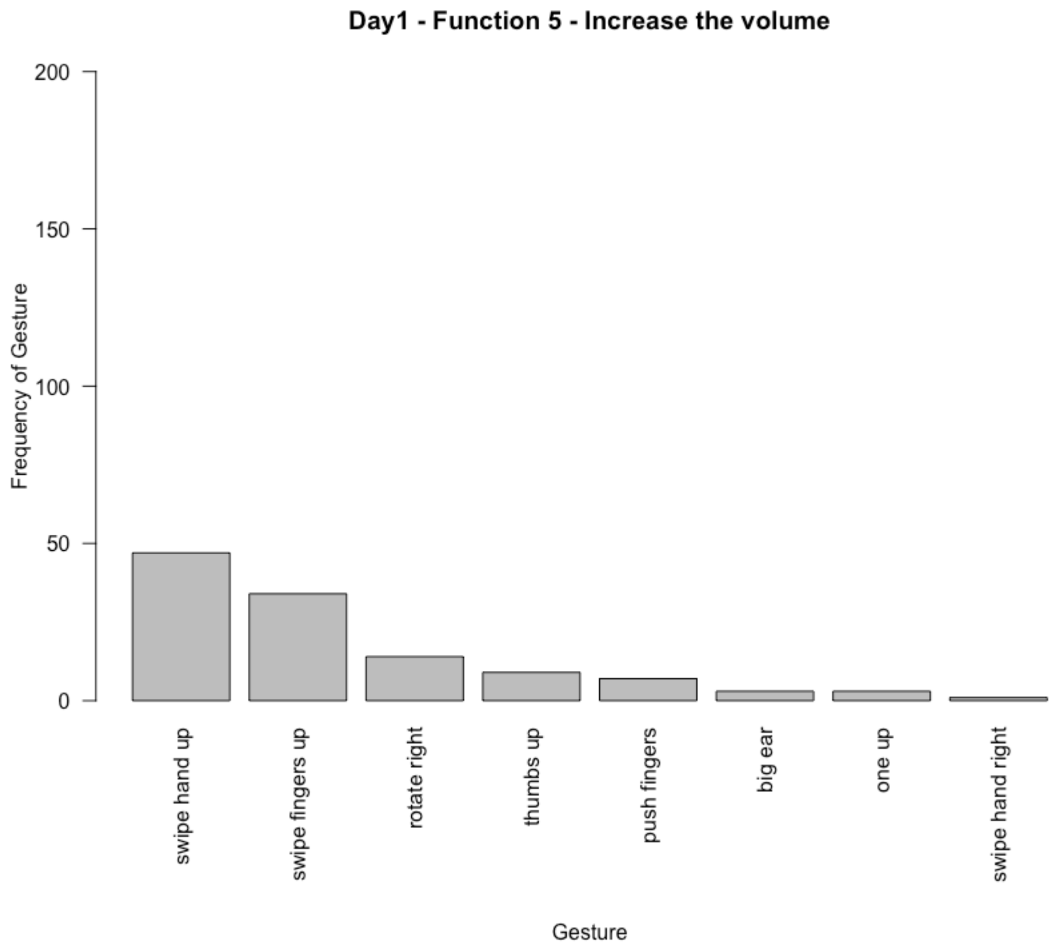
Appendix E – Day 1 bar graphs for intuitive gestures for controlling a presentation

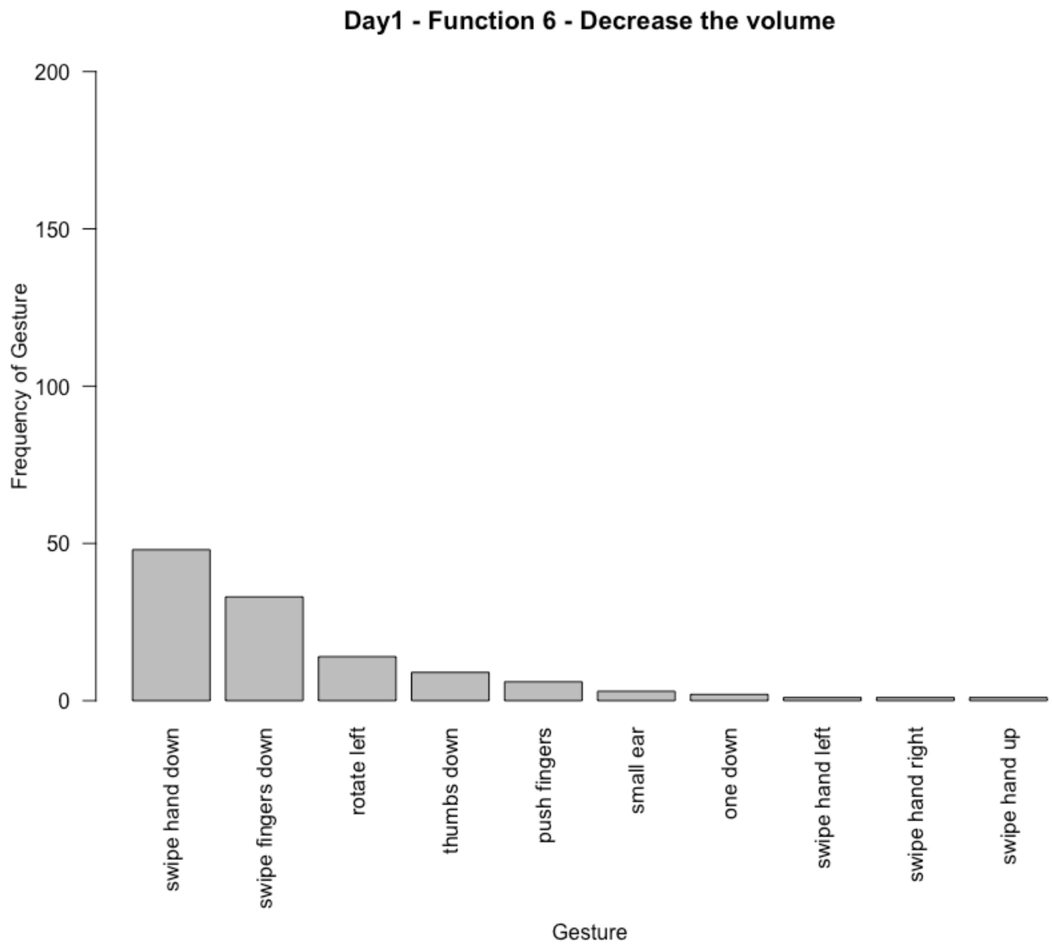



Day1 - Function 2 - Go to the previous slide
200
150
100
50
0
Frequency of Gesture
swipe hand right
swipe hand left
swipe fingers left
swipe fingers right
one left
swipe hand down
Gesture

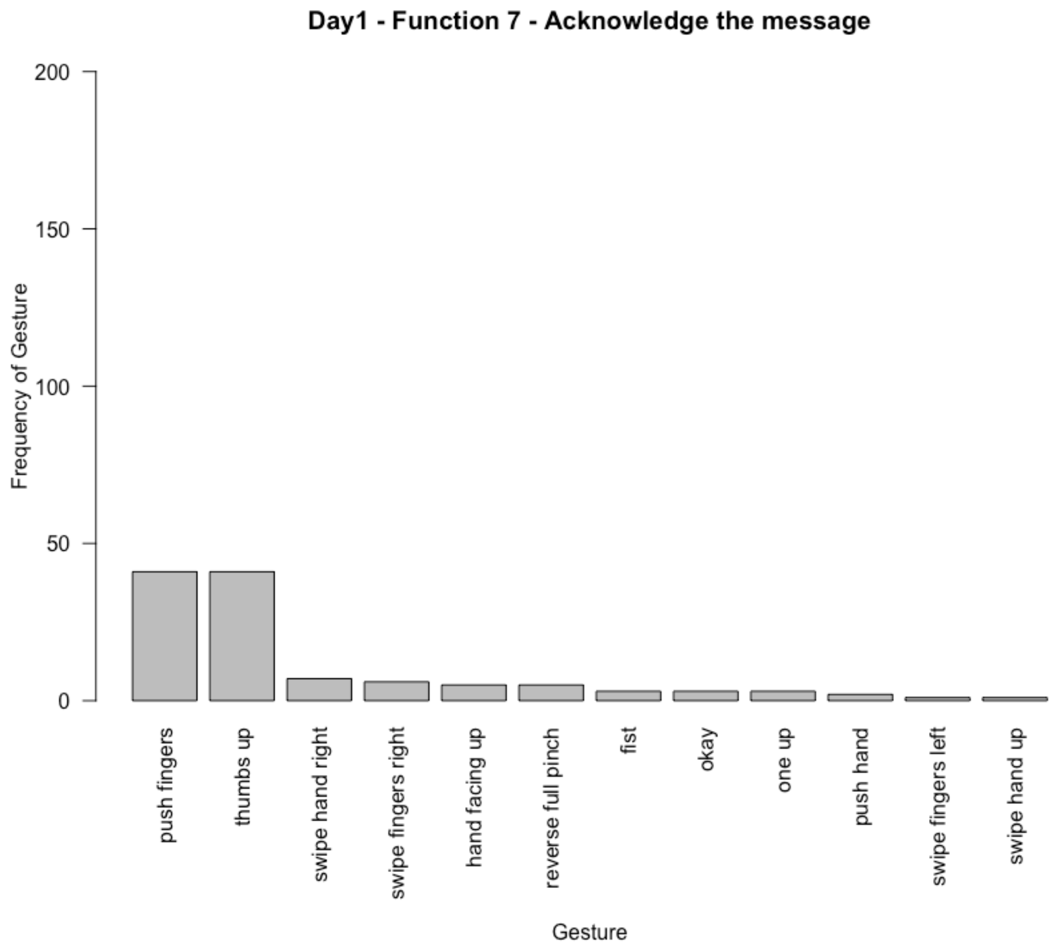

Day1 - Function 3 - Zoom in on the image
200
150
100
50
0
Frequency of Gesture
reverse full pinch
full pinch
swipe fingers left
push fingers
swipe hand up
Gesture

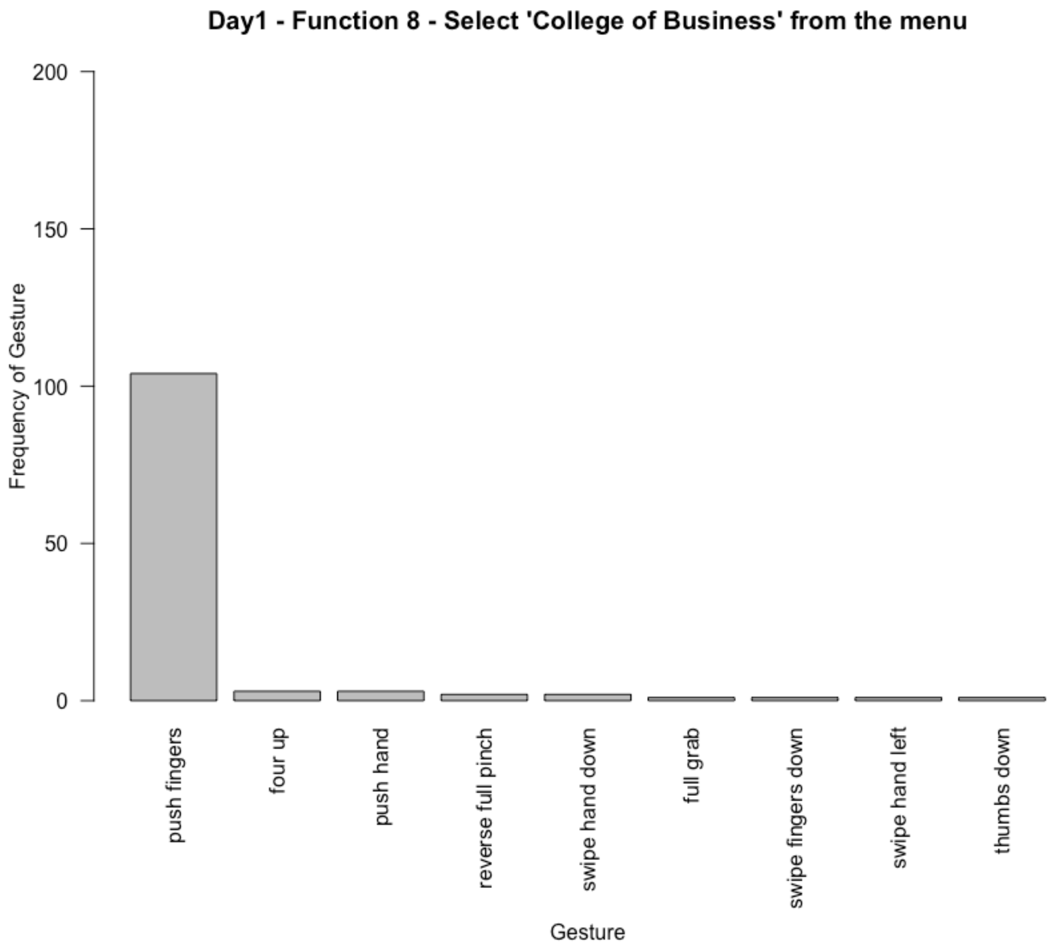

Day1 - Function 4 - Zoom out on the image
Frequency of Gesture
200
150
100
50
0
full pinch
reverse full pinch
push hand
swipe fingers left
swipe hand up
swipe hand right
Gesture

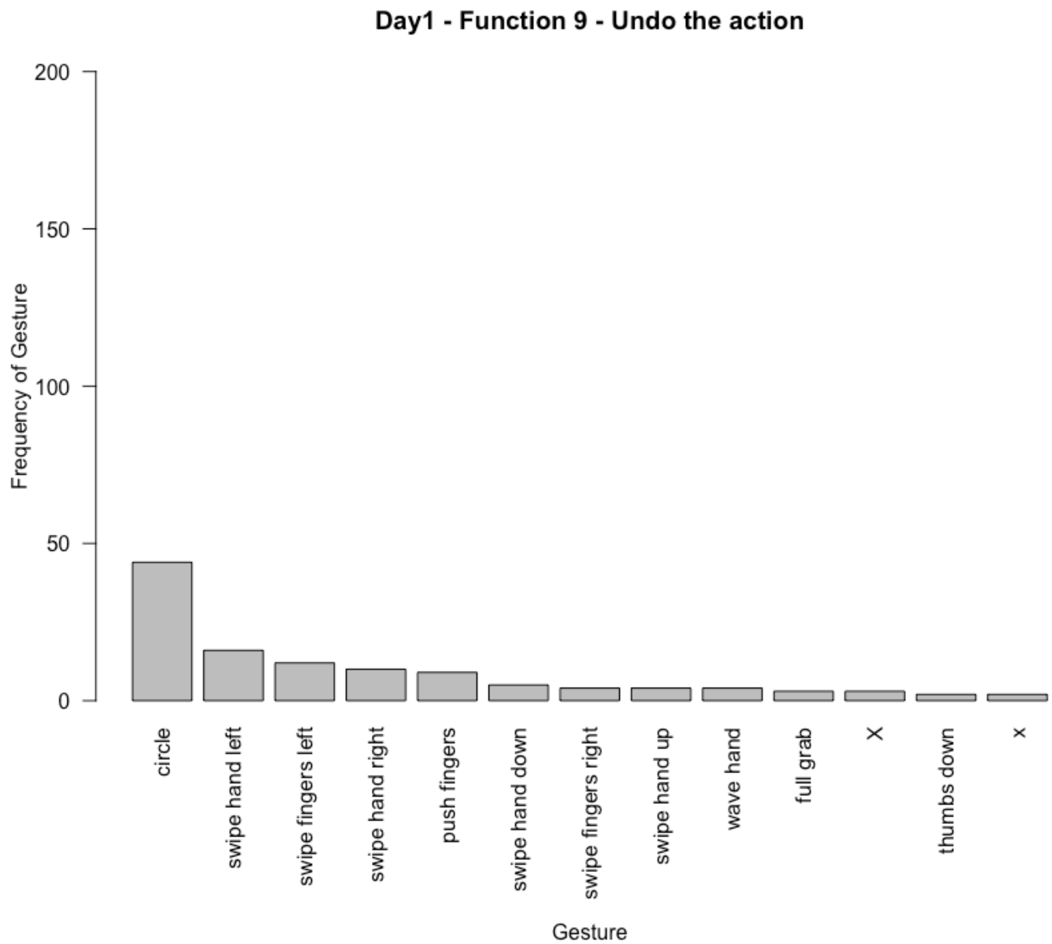

Day1 - Function 5 - Increase the volume
200
150
100
50
0
Frequency of Gesture
swipe hand up
swipe fingers up
rotate right
thumbs up
push fingers
big ear
one up
swipe hand right
Gesture



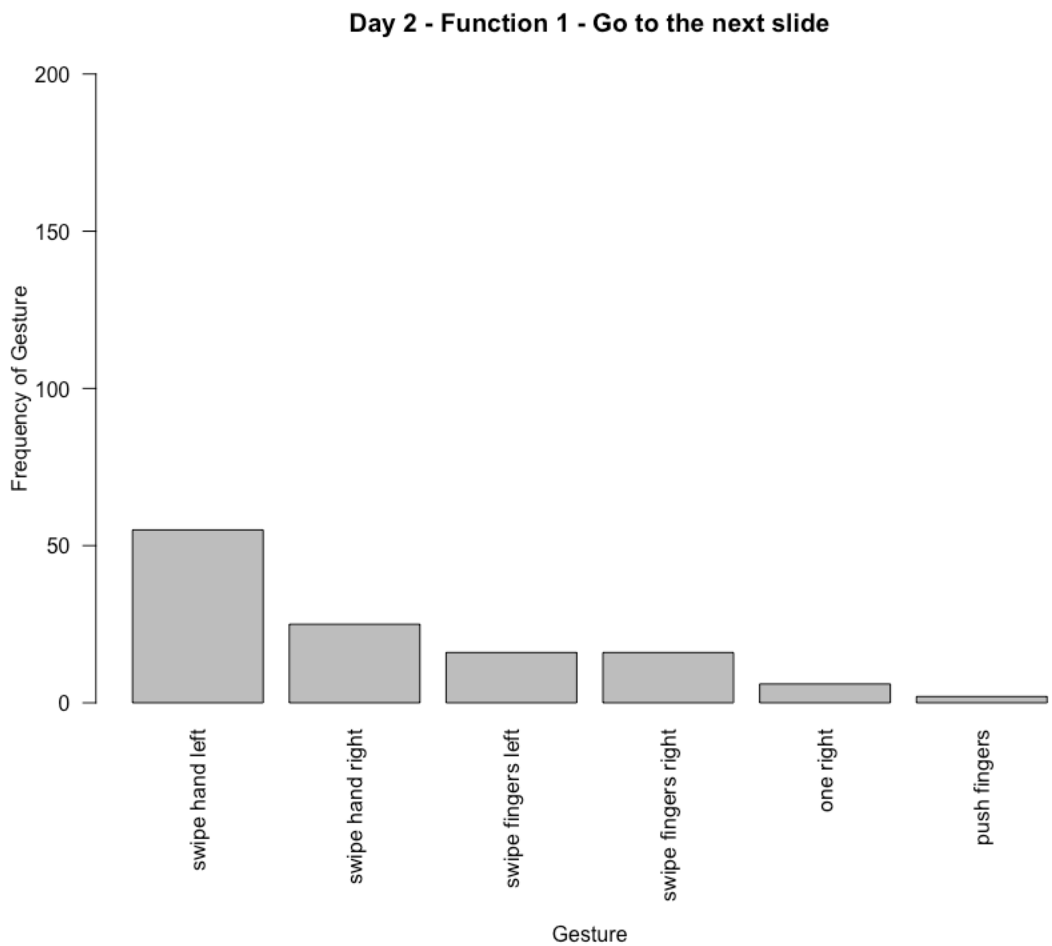
Day1 - Function 6 - Decrease the volume
Frequency of Gesture
200
150
100
50
0
swipe hand down
swipe fingers down
rotate left
thumbs down
push fingers
small ear
one down
swipe hand left
swipe hand right
swipe hand up
Gesture

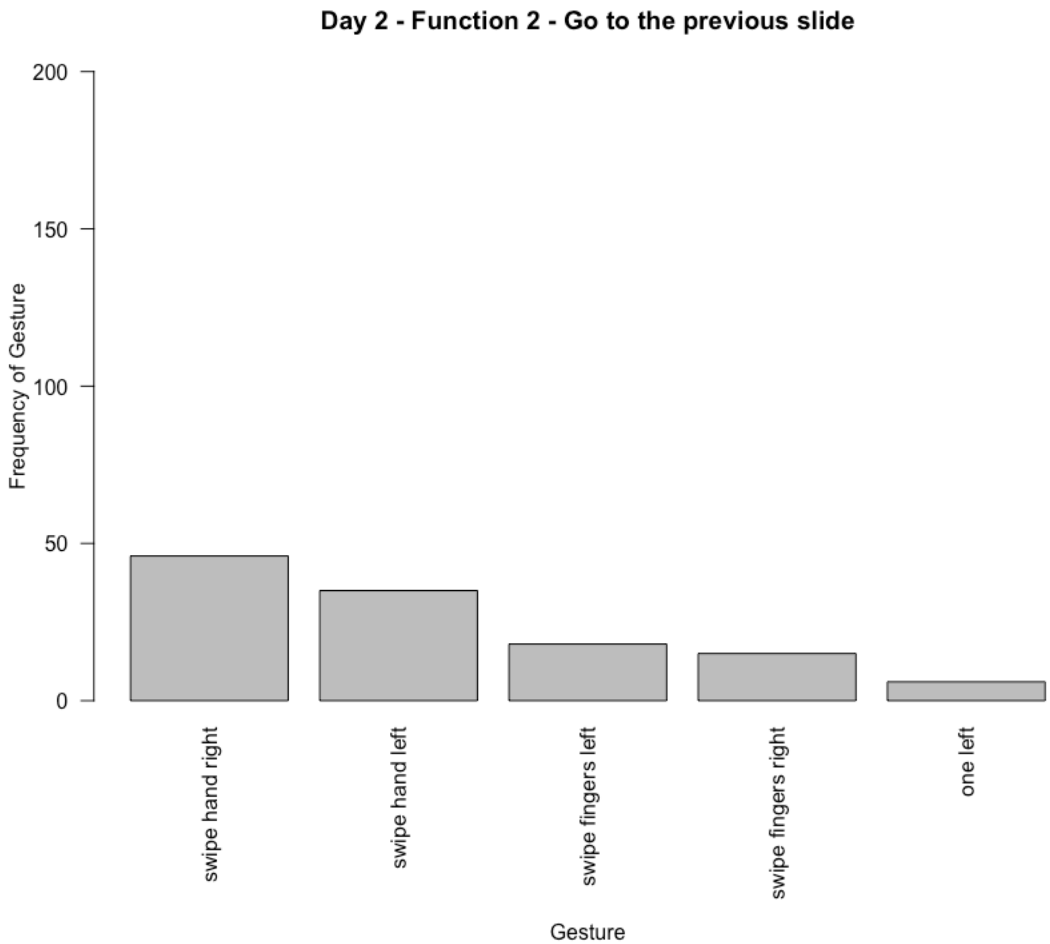

Day1 - Function 7 - Acknowledge the message
Frequency of Gesture
200
150
100
50
0
push fingers
thumbs up
swipe hand right
swipe fingers right
hand facing up
reverse full pinch
fist
okay
one up
push hand
swipe fingers left
swipe hand up
Gesture

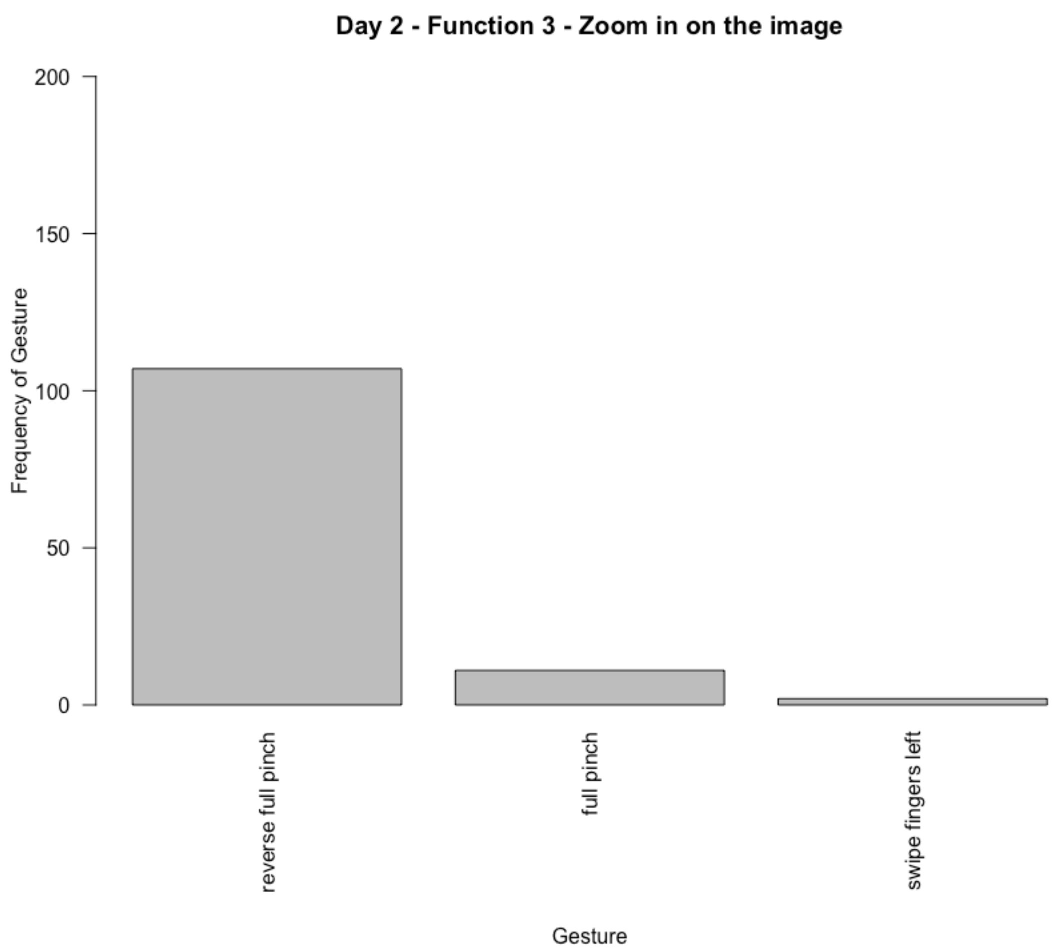

Day1 - Function 8 - Select 'College of Business' from the menu
Frequency of Gesture
200
150
100
50
0
push fingers
four up
push hand
reverse full pinch
swipe hand down
full grab
swipe fingers down
swipe hand left
thumbs down
Gesture

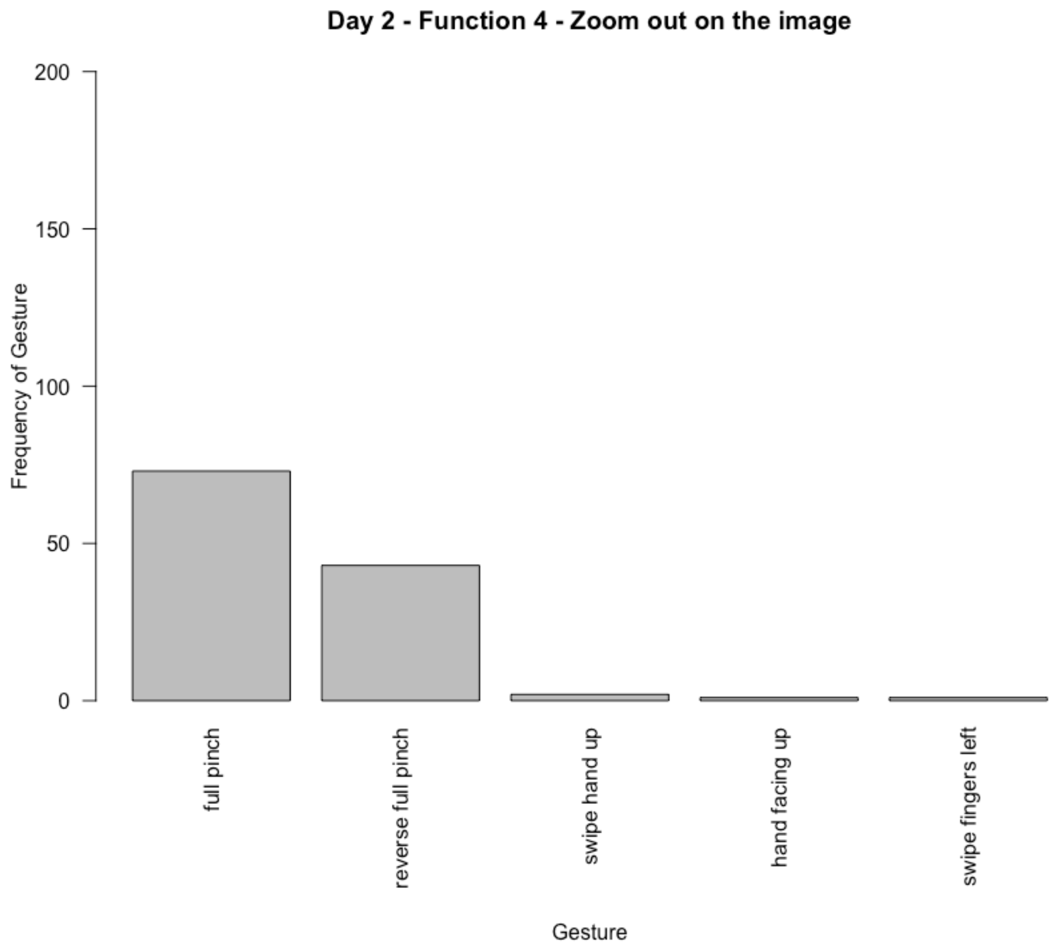

Day1 - Function 9 - Undo the action
200
150
100
50
0
Frequency of Gesture
circle
swipe hand left
swipe fingers left
swipe hand right
push fingers
swipe hand down
swipe fingers right
swipe hand up
wave hand
full grab
X
thumbs down
x
Gesture

Appendix F – Day 2 bar graphs for intuitive gestures for controlling a presentation

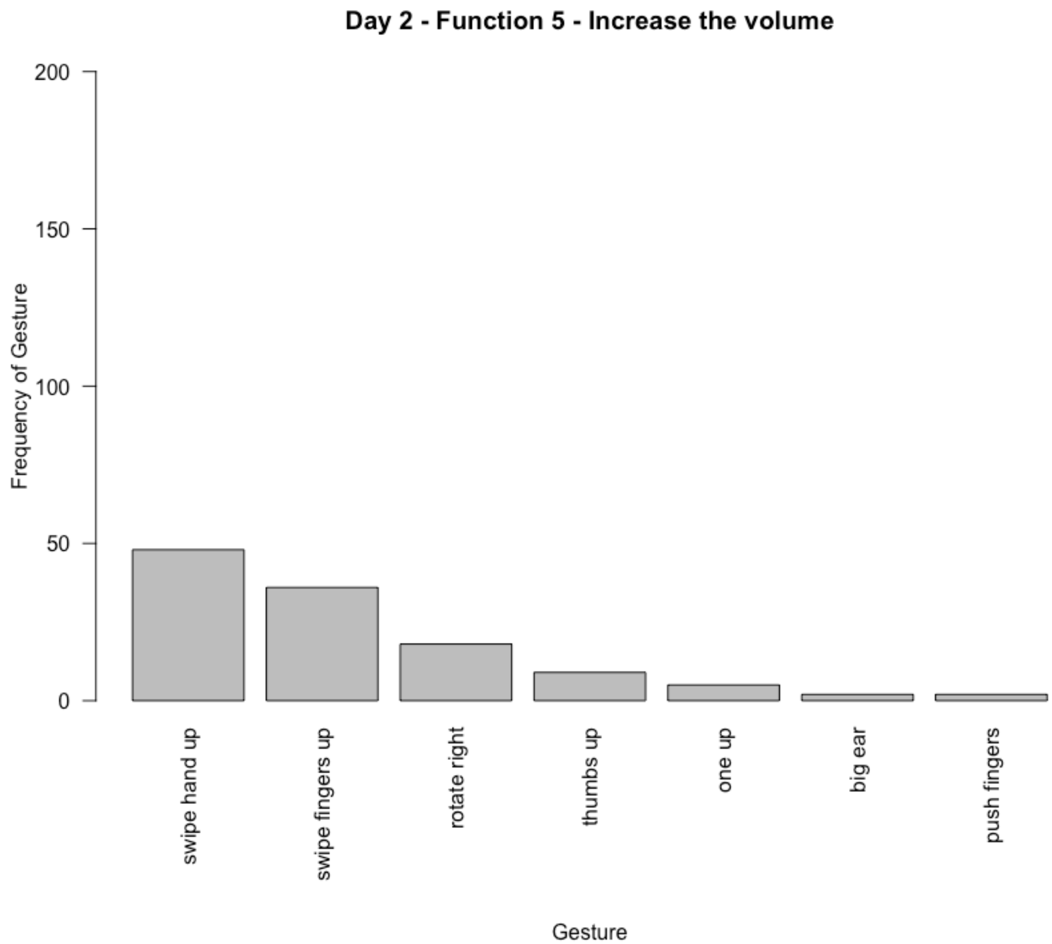



Day 2 - Function 2 - Go to the previous slide
Frequency of Gesture
200
150
100
50
0
swipe hand right
swipe hand left
swipe fingers left
swipe fingers right
one left
Gesture

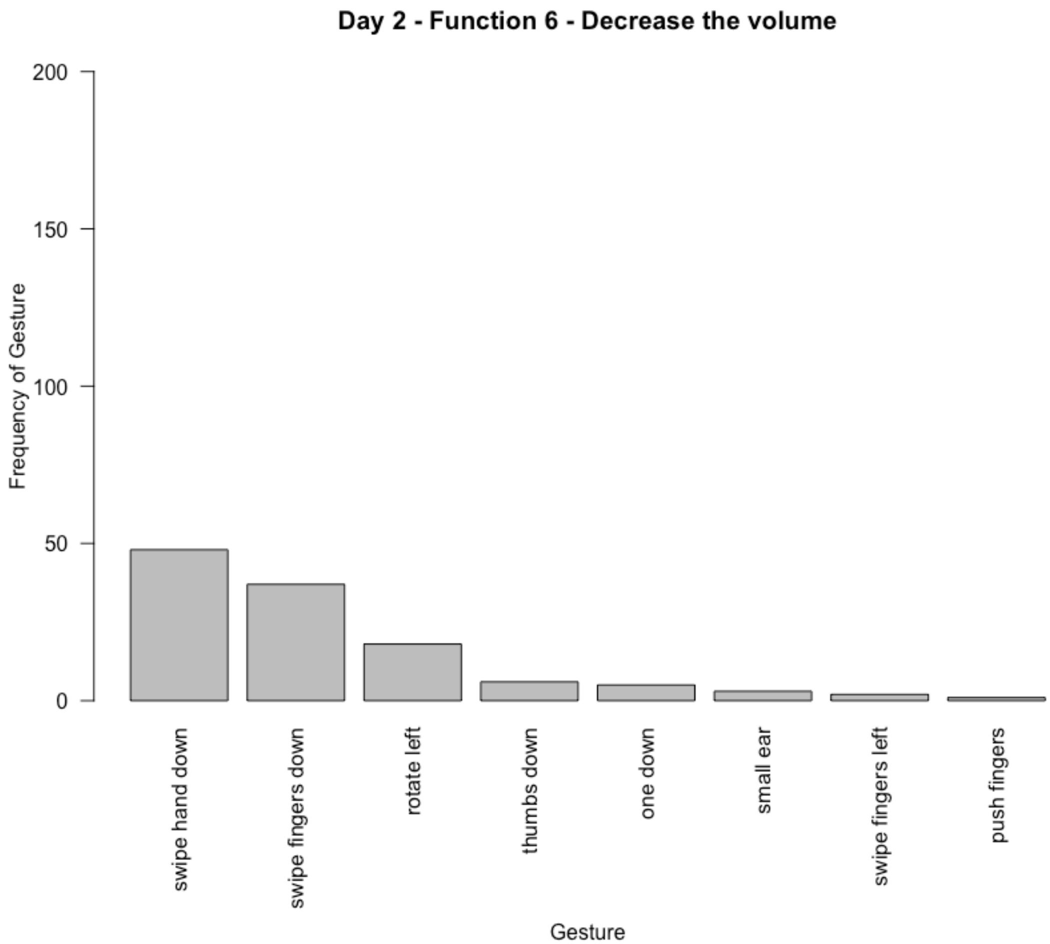

Day 2 - Function 3 - Zoom in on the image
200
150
100
50
0
Frequency of Gesture
reverse full pinch
full pinch
swipe fingers left
Gesture

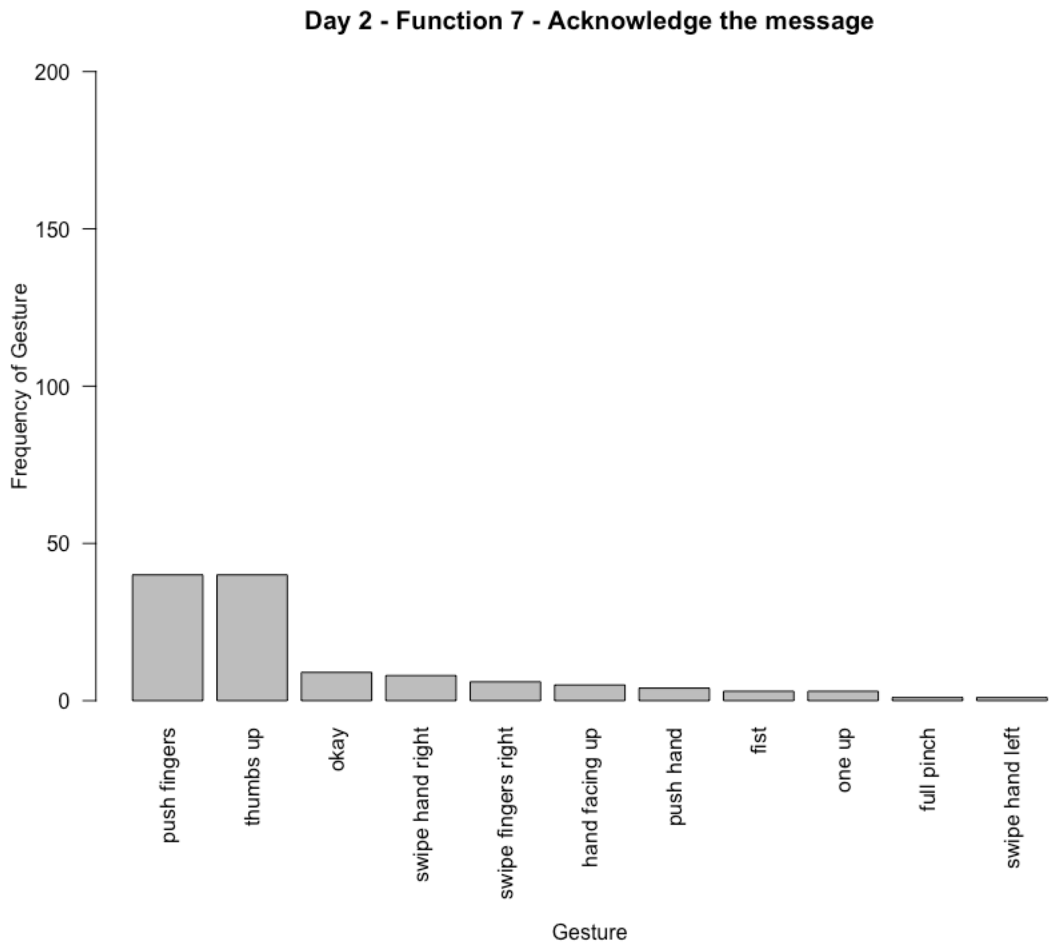


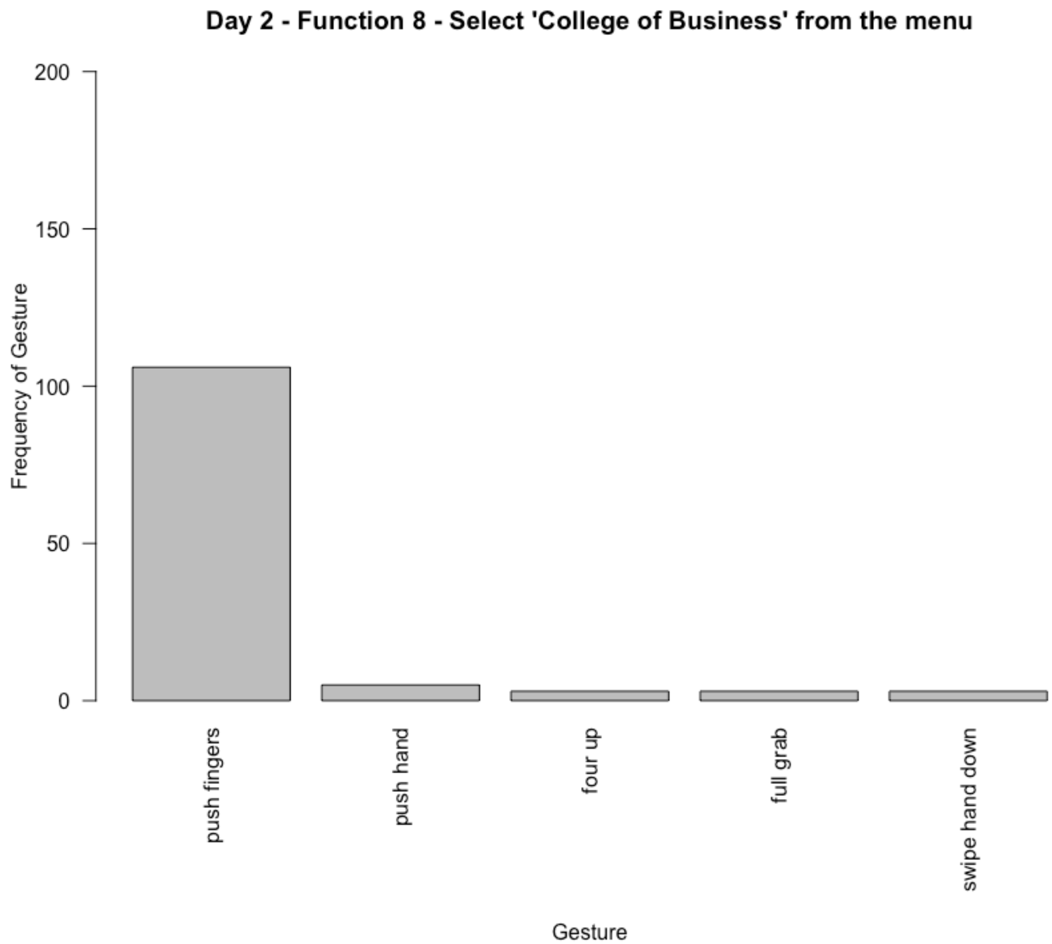
Day 2 - Function 4 - Zoom out on the image
200
150
100
50
0
Frequency of Gesture
full pinch
reverse full pinch
swipe hand up
hand facing up
swipe fingers left
Gesture

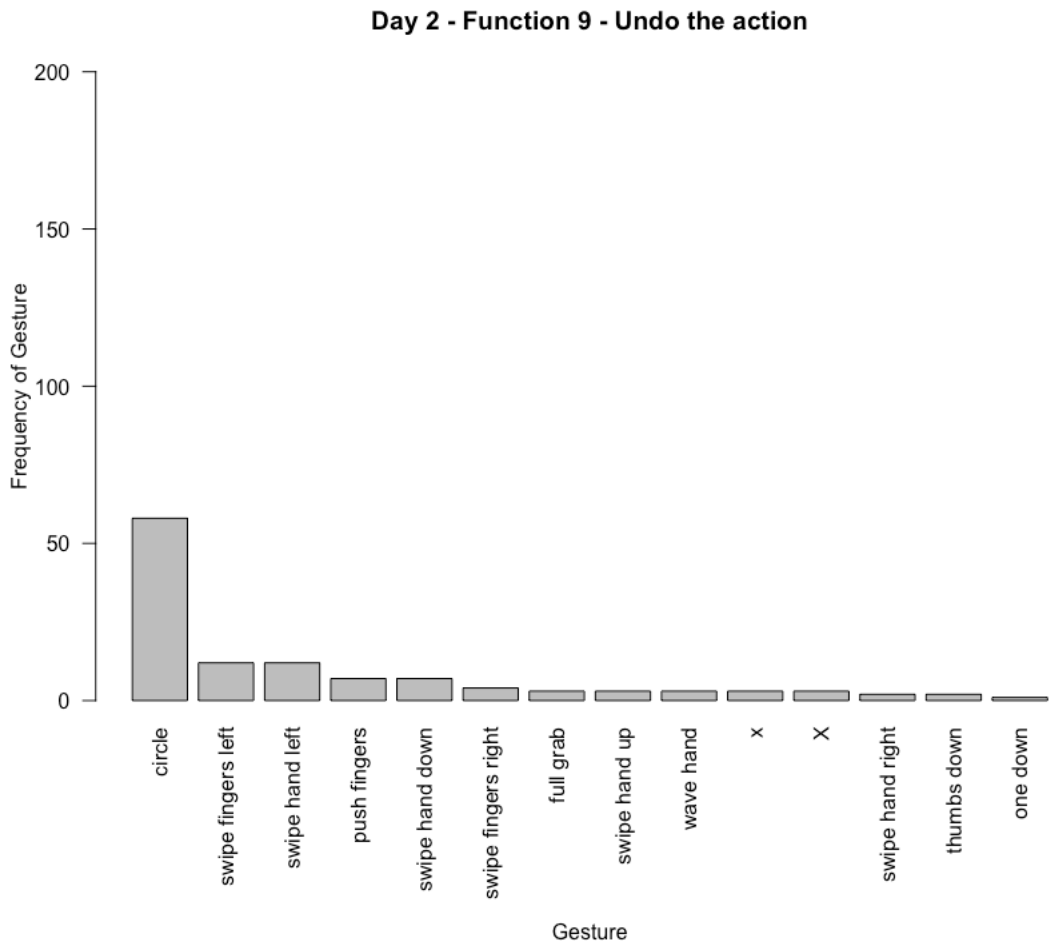


Day 2 - Function 5 - Increase the volume
Frequency of Gesture
200
150
100
50
0
swipe hand up
swipe fingers up
rotate right
thumbs up
one up
big ear
push fingers
Gesture

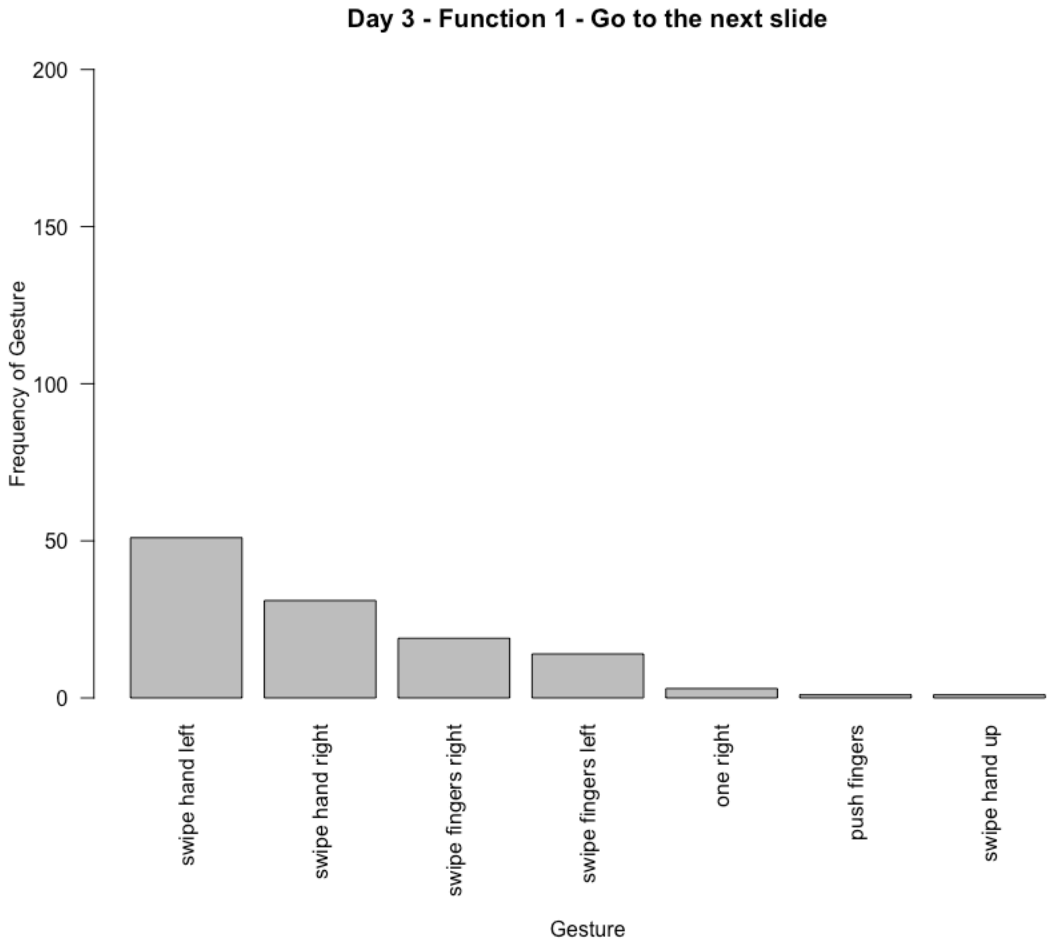

Day 2 - Function 6 - Decrease the volume
200
150
100
50
0
Frequency of Gesture
swipe hand down
swipe fingers down
rotate left
thumbs down
one down
small ear
swipe fingers left
push fingers
Gesture

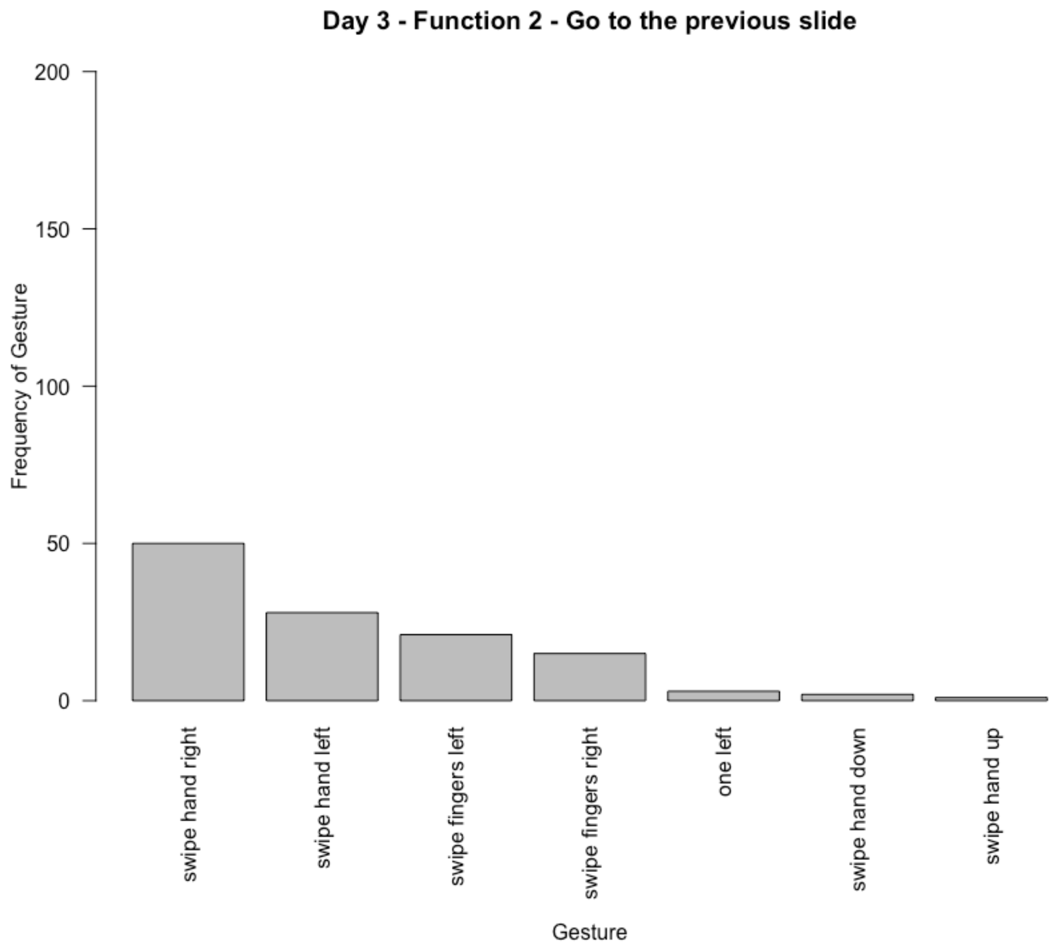

Day 2 - Function 7 - Acknowledge the message
200
150
100
50
0
Frequency of Gesture
push fingers
thumbs up
okay
swipe hand right
swipe fingers right
hand facing up
push hand
fist
one up
full pinch
swipe hand left
Gesture


Day 2 - Function 8 - Select 'College of Business' from the menu
200
150
100
50
0
Frequency of Gesture
push fingers
push hand
four up
full grab
swipe hand down
Gesture

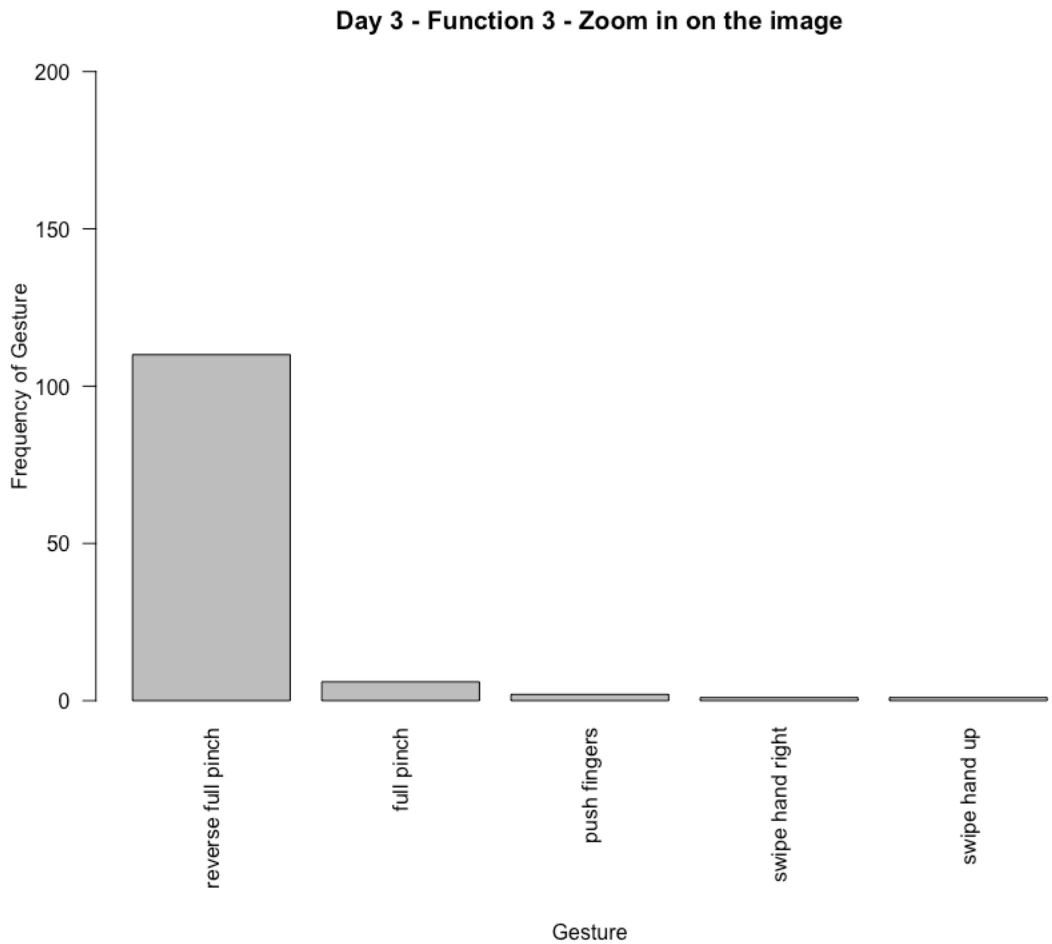

Day 2 - Function 9 - Undo the action
200
150
100
50
0
Frequency of Gesture
circle
swipe fingers left
swipe hand left
push fingers
swipe hand down
swipe fingers right
full grab
swipe hand up
wave hand
x
X
swipe hand right
thumbs down
one down
Gesture

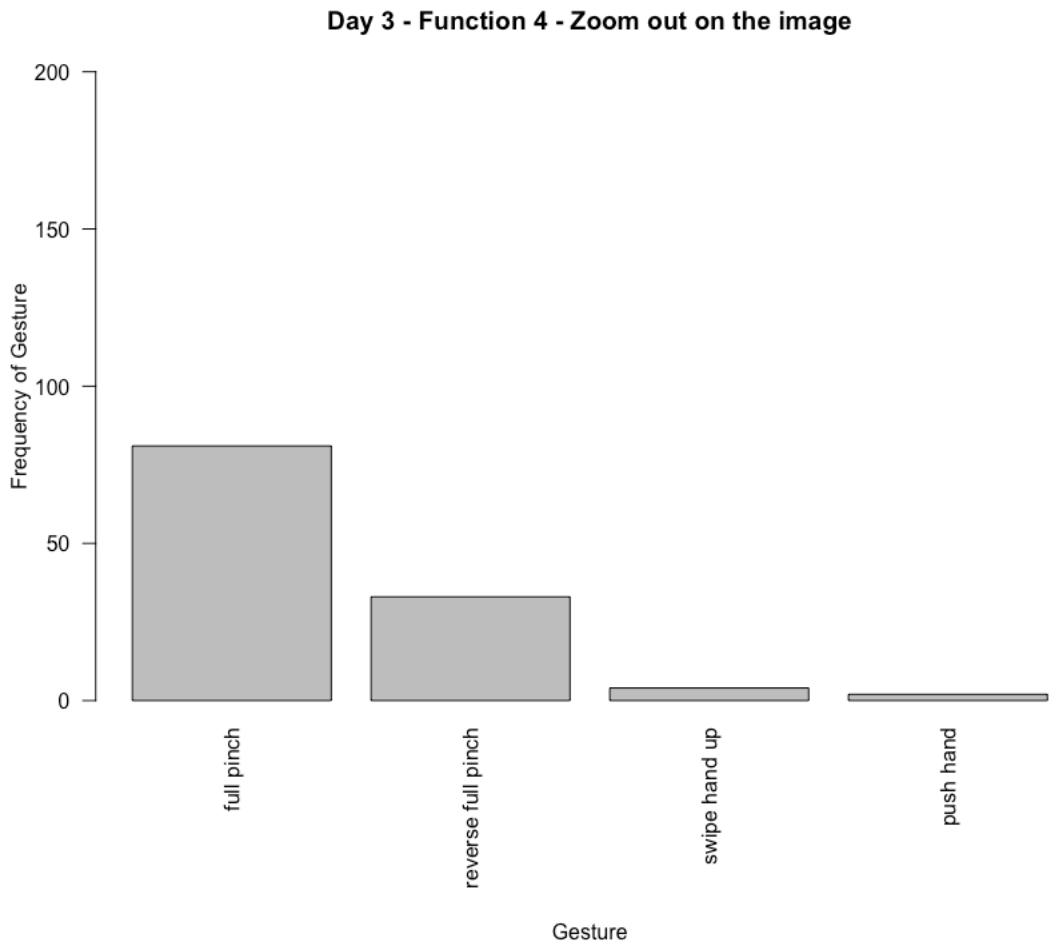
Appendix G – Day 3 bar graphs for intuitive gestures for controlling a presentation

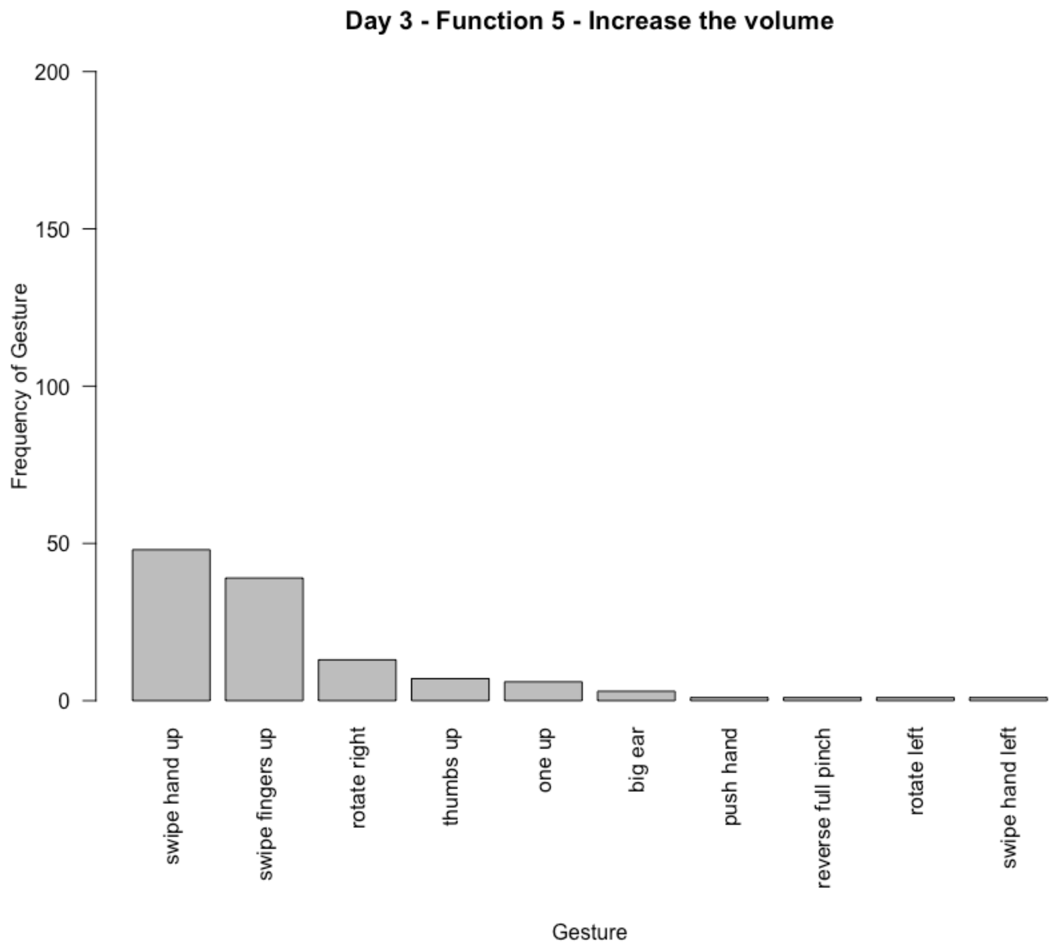


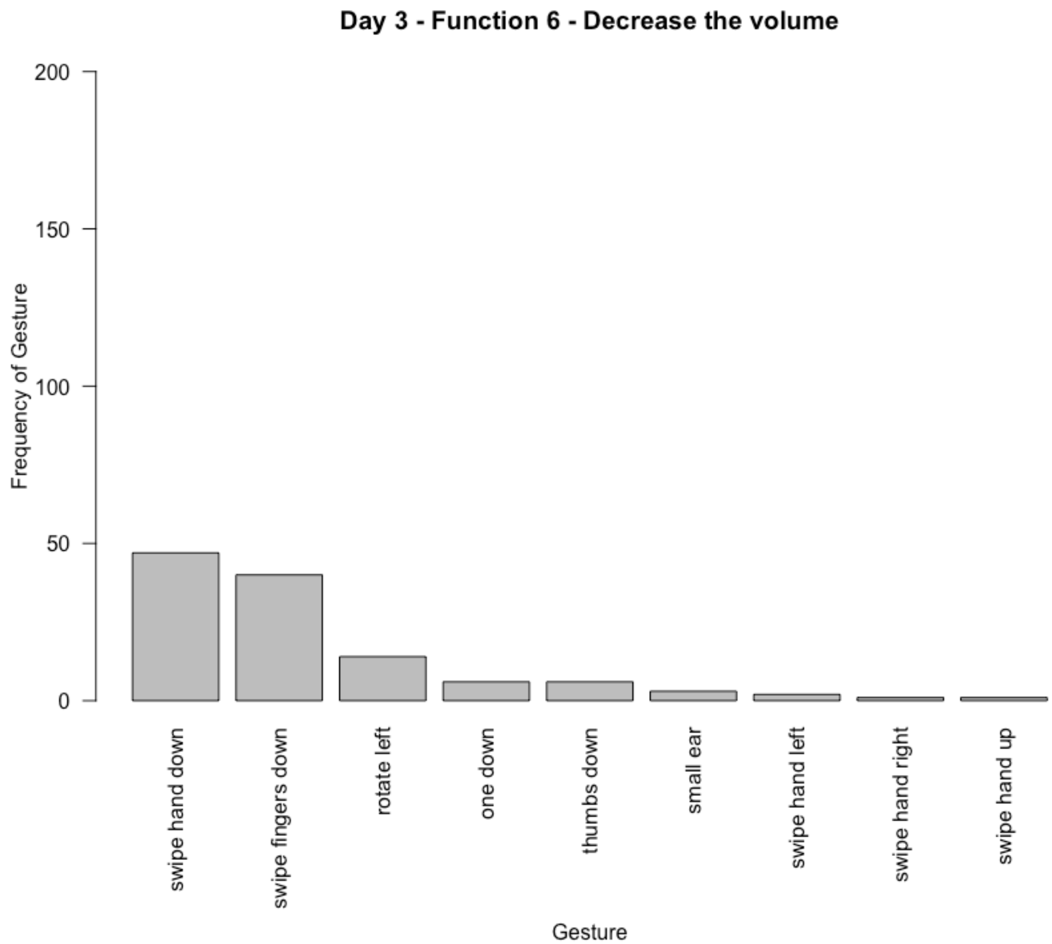


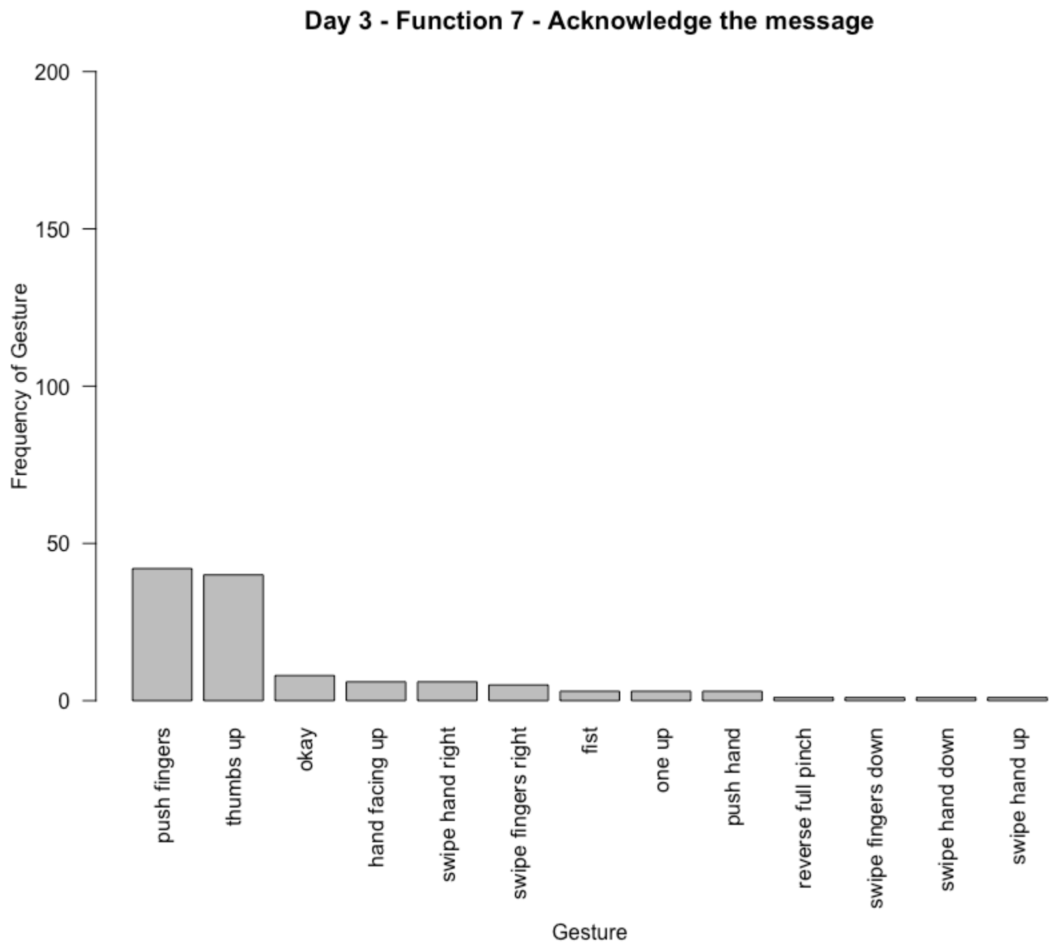
Day 3 - Function 2 - Go to the previous slide
200
150
100
50
0
Frequency of Gesture
swipe hand right
swipe hand left
swipe fingers left
swipe fingers right
one left
swipe hand down
swipe hand up
Gesture

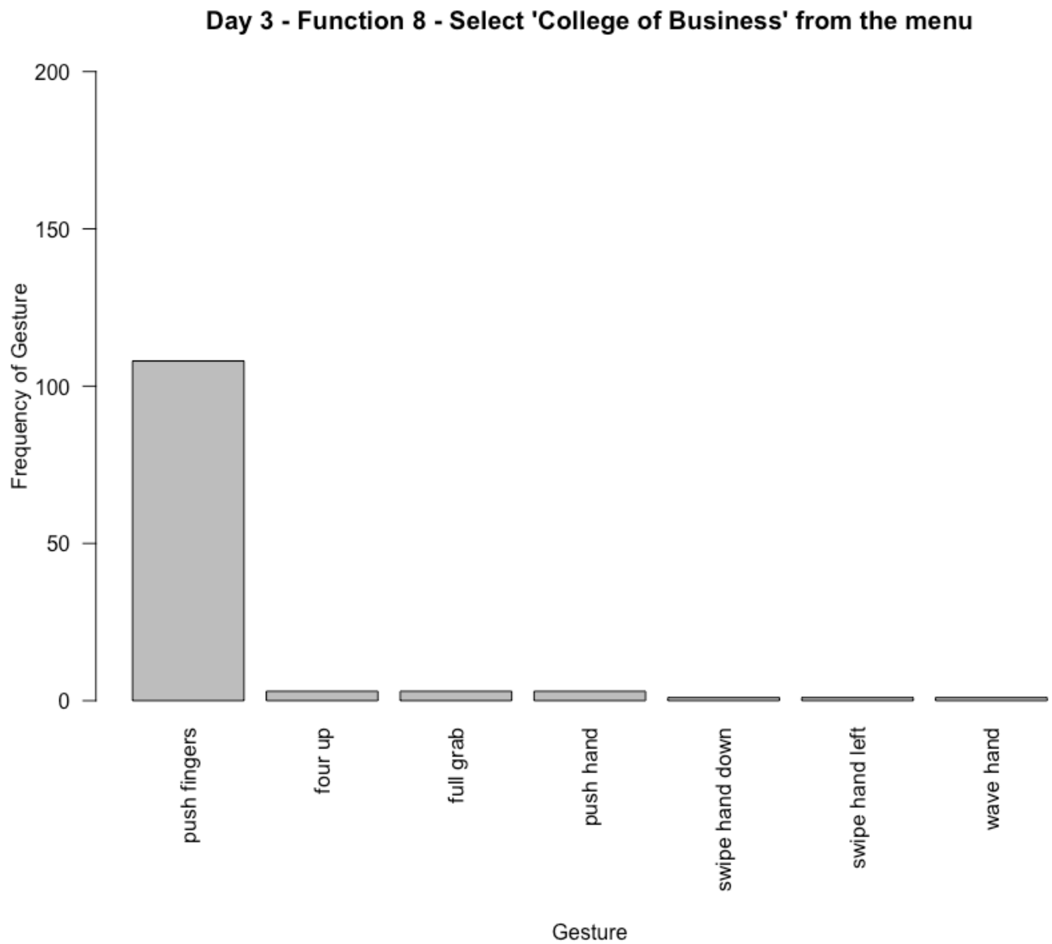

Day 3 - Function 3 - Zoom in on the image
200
150
100
50
0
Frequency of Gesture
reverse full pinch
full pinch
push fingers
swipe hand right
swipe hand up
Gesture

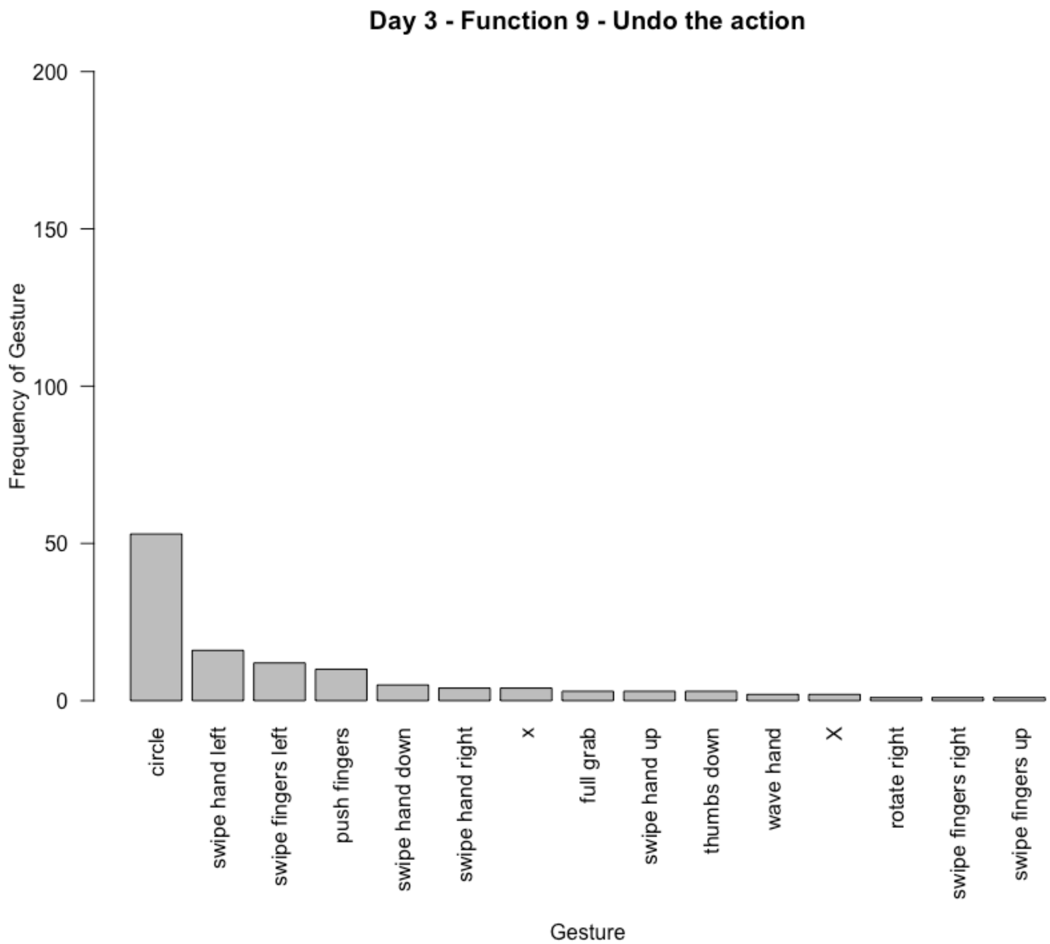

Day 3 - Function 4 - Zoom out on the image
Frequency of Gesture
0
50
100
150
200
full pinch
reverse full pinch
swipe hand up
push hand
Gesture


Day 3 - Function 5 - Increase the volume
200
150
100
50
0
Frequency of Gesture
swipe hand up
swipe fingers up
rotate right
thumbs up
one up
big ear
push hand
reverse full pinch
rotate left
swipe hand left
Gesture


Day 3 - Function 6 - Decrease the volume
Frequency of Gesture
200
150
100
50
0
swipe hand down
swipe fingers down
rotate left
one down
thumbs down
small ear
swipe hand left
swipe hand right
swipe hand up
Gesture


Day 3 - Function 7 - Acknowledge the message
200
150
100
50
0
Frequency of Gesture
push fingers
thumbs up
okay
hand facing up
swipe hand right
swipe fingers right
fist
one up
push hand
reverse full pinch
swipe fingers down
swipe hand down
swipe hand up
Gesture


Day 3 - Function 8 - Select 'College of Business' from the menu
Frequency of Gesture
200
150
100
50
0
push fingers
four up
full grab
push hand
swipe hand down
swipe hand left
wave hand
Gesture


Day 3 - Function 9 - Undo the action
200
150
100
50
0
Frequency of Gesture
circle
swipe hand left
swipe fingers left
push fingers
swipe hand down
swipe hand right
x
full grab
swipe hand up
thumbs down
wave hand
X
rotate right
swipe fingers right
swipe fingers up
Gesture

REFERENCES

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., … Koh, J. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI. *Microsoft Research TechReport MSR-TR-2012-111*.

Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000). Gesture Control for Use in Automobiles. In *MVA* (pp. 349–352).

Anderson, F., & Bischof, W. F. (2013). Learning and performance with gesture guides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1109–1118). ACM.

Ardito, C., Costabile, M. F., & Jetter, H.-C. (2014). Gestures that people can understand and use. *Journal of Visual Languages & Computing*, *25*(5), 572–576.

Ashley, S. (2014). Touch-less control coming to cars. *Automotive Engineering*, 21.

Barsuk, J. H., Cohen, E. R., Feinglass, J., McGaghie, W. C., & Wayne, D. B. (2009). Use of simulation-based education to reduce catheter-related bloodstream infections. *Archives of Internal Medicine*, *169*(15), 1420–1423.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, *36*(7), 28–35.

Betza, S. M., Jurewicz, K. A., Neyens, D. M., Riggs, S. L., Abernathy, J. H., & Reeves, S. T. (2016). Anesthesia Maintenance and Vigilance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, pp. 608–612). SAGE Publications. https://doi.org/10.1177/1541931213601139

Birnbach, D. J., Rosen, L. F., Fitzpatrick, M., Carling, P., & Munoz-Price, L. S. (2015). The Use of a Novel Technology to Study Dynamics of Pathogen Transmission in the Operating Room. *Anesthesia & Analgesia*, *120*(4).

Bizzotto, N., Costanzo, A., Bizzotto, L., Regis, D., Sandri, A., & Magnan, B. (2014). Leap motion gesture control with OsiriX in the operating room to control imaging: first experiences during live surgery. *Surgical Innovation*, *1*.

Cairnie, N., Ricketts, I. W., McKenna, S. J., & McAllister, G. (2000). Using finger-pointing to operate secondary controls in automobiles. In *Proceedings of the IEEE Intelligent Vehicles Symposium* (pp. 550–555). https://doi.org/10.1109/IVS.2000.898405

Choi, Y.-J. (2012). Exploring experiences of psychiatric nursing simulations using standardized patients for undergraduate students. *Asian Nursing Research*, *6*(3), 91–95.

Comer, A., Harris, A. D., Shardell, M., Braun, B., Belton, B. M., Wolfsthal, S. D., … Sulis, C. (2011). Web-based training improves knowledge about central line bloodstream infections. *Infection Control & Hospital Epidemiology*, *32*(12), 1219–1222.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15–18.

Cowles, M. K. (2013). *Applied Bayesian statistics: with R and OpenBUGS examples*

(Vol. 98). Springer Science & Business Media.
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.
Efron, D. (1941). *Gesture, race and culture: a tentative study of the spatio-temporal and" linguistic" aspects of the gestural behavior of eastern Jews and southern Italians in New York City, living under similar as well as different environmental conditions* (Vol. 9). Mouton.
Fariman, H. J., Alyamani, H. J., Kavakli, M., & Hamey, L. (2016). Designing a user-defined gesture vocabulary for an in-vehicle climate control system. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (pp. 391–395). ACM.
Freedman, N. (1972). The analysis of movement behavior during the clinical interview. *Studies in Dyadic Communication*, 153–175.
Freeman, D., Benko, H., Morris, M. R., & Wigdor, D. (2009). ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (pp. 165–172). ACM.
Freeman, W. T., & Weissman, C. (1995). Television control by hand gestures. In *Proc. of Intl. Workshop on Automatic Face and Gesture Recognition* (pp. 179–183).
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (2nd ed.). Chapman and Hall/CRC.
Henze, N., Löcken, A., Boll, S., Hesselmann, T., & Pielot, M. (2010). Free-hand gestures for music playback: deriving gestures with a user-centred process. In *Proceedings of the 9th international conference on Mobile and Ubiquitous Multimedia* (p. 16). ACM.
Hinckley, K., Jacob, R. J. K., Ware, C., Wobbrock, J. O., & Wigdor, D. (2014). Input/Output Devices and Interaction Techniques.
Horsky, J., Kaufman, D. R., Oppenheim, M. I., & Patel, V. L. (2003). A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *Journal of Biomedical Informatics*, *36*(1), 4–22.
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.
Höysniemi, J., Hämäläinen, P., & Turkki, L. (2004). Wizard of Oz Prototyping of Computer Vision Based Action Games for Children. In *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community* (pp. 27–34). New York, NY, USA: ACM.
Huang, J., Zhou, W., Li, H., & Li, W. (2015). Sign language recognition using real-sense. *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*. https://doi.org/10.1109/ChinaSIP.2015.7230384
Intel Corporation. (2016). Handling Gestures. *Intel® RealSense™ SDK 2016 R2 Documentation*.
Jacob, M. G., & Wachs, J. P. (2014a). Context-based hand gesture recognition for the operating room. *Pattern Recognition Letters*, *36*, 196–203.
Jacob, M. G., & Wachs, J. P. (2014b). *Context-based hand gesture recognition for the*

operating room. *Pattern Recognition Letters* (Vol. 36). https://doi.org/10.1016/j.patrec.2013.05.024

Jacob, M. G., Wachs, J. P., & Packer, R. A. (2013). Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *Journal of the American Medical Informatics Association*, *20*(1), 183–186.

Jæger, M. G., Skov, M. B., & Thomassen, N. G. (2008). You can touch, but you can't look: interacting with in-vehicle systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1139–1148). ACM.

Jurewicz, K. A., & Neyens, D. M. (2017). Mapping 3D Gestural Inputs to Traditional Touchscreen Interface Designs within the Context of Anesthesiology. In *Proceedings of the Human Factors and Ergonomics Society 61th Annual Meeting*. Austin, TX. https://doi.org/10.1177/1541931213601660

Jurewicz, K. A., Neyens, D. M., Catchpole, K., Joseph, A., Reeves, S. T., & Abernathy III, J. H. (n.d.). Using anaesthesia workflow to evaluate physical workspace design and layout. *British Journal of Anaesthesia*.

Jurewicz, K. A., Neyens, D. M., Catchpole, K., & Reeves, S. T. (2018). Developing a 3D gestural interface for anesthesia-related human-computer interaction tasks using both experts and novices. *Human Factors*, *60*(7), 992–1007. https://doi.org/10.1177/0018720818780544

Karam, M., & Schraefel, M. C. (2005). A Taxonomy of Gestures in Human Computer Interactions. University of Southampton. *Computing Service*.

Kendon, A. (1988). How gestures can become like words. *Cross-Cultural Perspectives in Nonverbal Communication*, *1*, 131–141.

Klauer, S. G., Guo, F., Sudweeks, J., & Dingus, T. A. (2010). *An analysis of driver inattention using a case-crossover approach on 100-car data*.

Klevens, R. M., Edwards, J. R., Richards, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., & Cardo, D. M. (2007). Estimating Health Care-Associated Infections and Deaths in U.S. Hospitals, 2002. *Public Health Reports (1974-)*, *122*(2), 160–166. Retrieved from http://www.jstor.org/stable/20057111

Loftus, R. W., Brown, J. R., Koff, M. D., Reddy, S., Heard, S. O., Patel, H. M., … Jensen, J. T. (2012). Multiple reservoirs contribute to intraoperative bacterial transmission. *Anesthesia & Analgesia*, *114*(6), 1236–1248.

Loftus, R. W., Muffly, M. K., Brown, J. R., Beach, M. L., Koff, M. D., Corwin, H. L., … Yeager, M. P. (2011). Hand contamination of anesthesia providers is an important risk factor for intraoperative bacterial transmission. *Anesthesia & Analgesia*, *112*(1), 98–105.

Marschall, J., Mermel, L. A., Fakih, M., Hadaway, L., Kallen, A., O'Grady, N. P., … Maragakis, L. L. (2014). Strategies to prevent central line-associated bloodstream infections in acute care hospitals: 2014 update. *Infection Control & Hospital Epidemiology*, *35*(S2), S89–S107.

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov chain monte carlo in R.

Mauney, D., Howarth, J., Wirtanen, A., & Capra, M. (2010). Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In *CHI'10*

Extended Abstracts on Human Factors in Computing Systems. ACM.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.
Mewes, A., Saalfeld, P., Riabikin, O., Skalej, M., & Hansen, C. (2016). A gesture-controlled projection display for CT-guided interventions. *International Journal of Computer Assisted Radiology and Surgery*, *11*(1), 157–164.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2015. *R Package Version*, 1–6.
Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010* (pp. 261–268). Canadian Information Processing Society.
Mortensen, D. (2017). Natural User Interfaces–What are they and how do you design user interfaces that feel natural. *Interaction Design Foundation*.
Nielsen, M., Störring, M., Moeslund, T., & Granum, E. (2004). A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction SE - 38* (Vol. 2915, pp. 409–420). Springer Berlin Heidelberg.
Nieuwenhuis, R., te Grotenhuis, H. F., & Pelzer, B. J. (2012). Influence. ME: tools for detecting influential data in mixed effects models.
Normark, C. J., Tretten, P., & Gärling, A. (2009). Do redundant head-up and head-down display configurations cause distractions? *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design*, 398–404.
Ohn-Bar, E., & Trivedi, M. M. (2014). Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations. *IEEE Transactions on Intelligent Transportation Systems*, *15*(6), 2368–2377. https://doi.org/10.1109/TITS.2014.2337331
Parada-Loira, F., González-Agulla, E., & Alba-Castro, J. L. (2014). Hand gestures to control infotainment equipment in cars. In *2014 IEEE Intelligent Vehicles Symposium Proceedings* (pp. 1–6). https://doi.org/10.1109/IVS.2014.6856614
Pereira, A., Wachs, J. P., Park, K., & Rempel, D. (2015). A user-developed 3-D hand gesture set for human–computer interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(4), 607–621.
Pickering, C. A. (2005). The search for a safer driver interface: a review of gesture recognition human machine interface. *Computing & Control Engineering Journal*, *16*(1), 34–40. https://doi.org/10.1049/cce:20050109
Pickering, C. A., Burnham, K. J., & Richardson, M. J. (2007). A research study of hand gesture recognition technologies and applications for human vehicle interaction. In *3rd Conference on Automotive Electronics*.
Pisharady, P. K., & Saerbeck, M. (2015). Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*,

141, 152–165. https://doi.org/https://doi.org/10.1016/j.cviu.2015.08.004
Pittet, D., Hugonnet, S., Harbarth, S., Mourouga, P., Sauvan, V., Touveneau, S., & Perneger, T. V. (2000). Effectiveness of a hospital-wide programme to improve compliance with hand hygiene. *The Lancet*, *356*(9238), 1307–1312. https://doi.org/http://dx.doi.org/10.1016/S0140-6736(00)02814-2
Rahman, A. S. M., Saboune, J., & El Saddik, A. (2011). Motion-path based in car gesture control of the multimedia devices. In *Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications* (pp. 69–76). ACM.
Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(3), 257–266. https://doi.org/10.1109/TSMC.1983.6313160
Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, *43*(1), 1–54. https://doi.org/10.1007/s10462-012-9356-9
Riener, A. (2012). Gestural Interaction in Vehicular Applications. *Computer*, *45*(4), 42–47. https://doi.org/10.1109/MC.2012.108
Riener, A., Ferscha, A., Bachmair, F., Hagmüller, P., Lemme, A., Muttenthaler, D., … Weger, F. (2013). Standardization of the in-car gesture interaction space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 14–21). ACM.
Riener, A., & Rossbory, M. (2011). Natural and intuitive hand gestures: a substitute for traditional vehicle control? *Proceedings of the AutomotiveUI*, *11*, 2.
Robert, C. P., Elvira, V., Tawn, N., & Wu, C. (2018). Accelerating MCMC algorithms. *Wiley Interdisciplinary Reviews: Computational Statistics*, *10*(5), e1435.
Sax, H., Allegranzi, B., Uçkay, I., Larson, E., Boyce, J., & Pittet, D. (2007). 'My five moments for hand hygiene': a user-centred design approach to understand, train, monitor and report hand hygiene. *Journal of Hospital Infection*, *67*(1), 9–21. https://doi.org/http://dx.doi.org/10.1016/j.jhin.2007.06.004
Schroder, S., Loftfield, N., Langmann, B., Frank, K., & Reithmeier, E. (2014). Contactless operating table control based on 3D image processing. *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. https://doi.org/10.1109/EMBC.2014.6943610
Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced" complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, *3*(2), 111–122.
Stackhouse, R. A., Beers, R., Brown, D., Brown, M., Greene, E., McCann, M. E., & Tompkins, B. M. (2011). *Recommendations for Infection Control for the Practice of Anesthesiology*. *American Society of Anesthesiologists* (3rd ed.). Park Ridge, IL.
Stern, H. I., Wachs, J. P., & Edan, Y. (2006). Human Factors for Design of Hand Gesture Human - Machine Interaction. *IEEE International Conference on Systems Man and Cybernetics*, *5*, 4052–4056. Retrieved from http://web.ics.purdue.edu/~jpwachs/papers/smc06_human.pdf

Stern, H. I., Wachs, J. P., & Edan, Y. (2008). Optimal Consensus Intuitive Hand Gesture Vocabulary Design. *2008 IEEE International Conference on Semantic Computing*. https://doi.org/10.1109/ICSC.2008.29

Sturman, D. J., & Zeltzer, D. (1994). A survey of glove-based input. *IEEE Computer Graphics and Applications*, *14*(1), 30–39.

The Society of Healthcare Epidemiology of America. (2010). Enhancing Patient Safety by Reducing Healthcare-Associated Infections: The Role of Discovery and Dissemination. *Infection Control and Hospital Epidemiology : The Official Journal of the Society of Hospital Epidemiologists of America*, *31*(2), 118–123. https://doi.org/10.1086/650198

U.S. Department of Health and Human Services. (2013). National Action Plan to prevent health care-associated infections: road map to elimination. *Washington, DC: USDHHS*. Retrieved from http://www.health.gov/hai/pdfs/hai-action-plan-executive-summary.pdf

Uva, A. E., Fiorentino, M., Manghisi, V. M., Boccaccio, A., Debernardis, S., Gattullo, M., & Monno, G. (2019). A User-Centered Framework for Designing Midair Gesture Interfaces. *IEEE Transactions on Human-Machine Systems*, *49*(5), 421–429.

Van der Meer, T., Te Grotenhuis, M., & Pelzer, B. (2010). Influential cases in multilevel modeling a methodological comment. *American Sociological Review*, *75*(1), 173–178.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(4), 589–606. https://doi.org/10.1109/21.156574

Wachs, J. P., Kölsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, *54*(2), 60–71.

Wachs, J. P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C., & Smith, M. (2008). A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, *15*(3), 321–323.

Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., & Smith, M. (2006). A real-time hand gesture interface for a medical image guided system. In *In ISRACAS 2006-Ninth Israeli Symposium on Computer-Aided Surgery, Medical Robotics, and Medical Imaging*. Citeseer.

Weber, D. J., Anderson, D., & Rutala, W. A. (2013). The role of the surface environment in healthcare-associated infections. *Current Opinion in Infectious Diseases*, *26*(4), 338–344.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. Retrieved from http://ggplot2.org

Wigdor, D., & Wixon, D. (2011). *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1083–1092). ACM.

World Health Organization. (2009). *WHO guidelines on hand hygiene in health care:*

first global patient safety challenge. Clean care is safer care. World Health Organization.

Yale, S. H., Hansotia, P., Knapp, D., & Ehrfurth, J. (2003). Neurologic conditions: assessing medical fitness to drive. *Clinical Medicine & Research*, *1*(3), 177–188.