Teaching users how to interact with gesture-based interfaces; a comparison of teaching-methods

By Eline Jansen

Student identity number: 0580766

in partial fulfilment of the requirements for the degree of

Master of Science

in Human-Technology Interaction

Supervisors:
Dr. Ir. Raymond Cuijpers       TU/e, IEIS, Human-Technology Interaction
Dr. Wijnand IJsselsteijn       TU/e, IEIS, Human-Technology Interaction
Ir. Ing. Maurice van Beurden   TU/e, IEIS, Human-Technology Interaction
Ir. John de Vet                Philips Consumer Lifestyle, BG TV Site Eindhoven
Keywords: gesture, gesture based, intuitive interaction, television, learning methods, teaching methods, human-technology
Preface

This thesis is my final assignment for the Master program Human-Technology Interaction at the University of Technology Eindhoven. Like all parts of the master program, it required a lot of effort, but I enjoyed working on every part of it. I would like to thank all staff and students of the HTI department for this experience. Specially, I would like to thank my supervisors Raymond Cuijpers, Wijnand IJsselsteijn, and Maurice van Beurden of the HTI-group for their support, enthusiasm, and sometimes patients. In addition, I would like to thank John the Vet for the opportunity to perform this graduation project in cooperation with the BG TV Site Eindhoven of Philips Consumer Lifestyle.
Summary

This graduation project aims to investigate two topics in the field of gesture-based interfaces: intuitive interaction gestures for controlling a television and teaching-methods for teaching users gesture-sets. Two experiments have been performed. The first experiment explored which television-interaction gestures were considered intuitive by means of an improvisation-task performed by fifteen participants. The results showed many commonalities between improvised gestures of participants, especially for interaction tasks that were perceived as easy. Most intuitive gestures for easy tasks were pantomimes, whereas for more difficult tasks most intuitive gestures were emblems.

Ten of the intuitive gestures that resulted from the first experiment were used in the gesture-set of the second experiment. In this experiment three teaching methods were compared by teaching a gesture-set to 120 participants, by one of three methods. In the first method, called imitation, participants observed a human arm performing the gesture. In the second method, called pantomime, participants observed an animated interface-object that communicates the gesture by means of object affordances. And the third method was a combination method in which the participant simultaneously observed both stimuli. The gestures were categorized as either intuitive or unintuitive (depending on the inherent task-association), and half of the participants were instructed to practice the movement while the other half was instructed not to practice. To control for the potential effect of gesture type both pantomime gestures and emblems were tested. The dependent variables were the percentage of correct recall and the response time of participants and were measured during four recall tests. The effects on the dependent variables were divided in learnability and memorability effects. Learnability effects were measured during the first three tests. Memorability effects were measured in the fourth test that was subsequent to a distraction task.

The results of the second experiment showed that all three teaching-methods resulted in good recall for gestures. After three observations participants correctly recalled, on average, eight gestures. The results showed a main effect for teaching-method and intuitiveness, but not for practice. In addition, the results showed teaching-method interaction-effects for intuitiveness, practice, and gesture type. The results of the pantomime teaching-method in which participants were only exposed to the goal of the movement, were influenced most by the interaction effects. Recall was least affected for conditions in which both the means (movement) and the goal were
communicated. This suggests that although humans are mainly interested in a movement’s goal during observational learning, the observed movements of gestures are processed and are assumed to complement memory processes significantly.
Table of contents

Preface .......................................................................................................................... II
Summary ......................................................................................................................... IV

1. Introduction ................................................................................................................... 1
2. Experiment 1 ................................................................................................................... 13
   2.1 Method ..................................................................................................................... 13
       Design ....................................................................................................................... 13
       Participants ............................................................................................................... 13
       Apparatus & Set-up ................................................................................................. 14
       Procedure ................................................................................................................ 14
   2.2 Results ..................................................................................................................... 15
       Commonalities .......................................................................................................... 15
       Association types ..................................................................................................... 28
       Scenario differences ............................................................................................... 29
   2.3 Discussion ............................................................................................................... 31
3. Experiment 2 ................................................................................................................ 32
   3.1 Method ..................................................................................................................... 32
       Design ....................................................................................................................... 32
       Participants ............................................................................................................... 33
       Set-up & apparatus ................................................................................................. 33
       Procedure ................................................................................................................ 34
       Annotation & data analysis ..................................................................................... 35
   3.2 Results ..................................................................................................................... 39
       Main effects .............................................................................................................. 39
       Interaction effects .................................................................................................. 42
4. General discussion ...................................................................................................... 45
   Implications for design ............................................................................................... 50

5. References .................................................................................................................. 53
Appendices experiment 1:

Appendix A. Description scenarios ........................................................................................................... 57
Appendix B. Questionnaire.......................................................................................................................... 58
Appendix C. General instruction & Informed consent ................................................................................. 60
Appendix D. Set-up experiment ................................................................................................................... 61
Appendix E. Description variables data-file ............................................................................................... 62
Appendix F. Gesture commonalities............................................................................................................. 65

Appendices preliminary tests:

Appendix G. Preliminary tests .................................................................................................................... 66
Appendix H. Informed consent preliminary tests ......................................................................................... 83
Appendix I. Instructions preliminary tests ................................................................................................. 84

Appendices experiment 2:

Appendix J. Overview stimuli ..................................................................................................................... 85
Appendix K. Criteria for recall judgments ................................................................................................... 89
Appendix L. Stimuli characteristics ............................................................................................................. 90
Appendix M. Set-up experiment ................................................................................................................... 92
Appendix N. Questionnaire & distraction task ............................................................................................ 93
Appendix O. General instruction & informed consent ................................................................................ 96
Appendix P. Instructions experiment ......................................................................................................... 97
Appendix Q. Logistic regression recall models ........................................................................................... 100

Appendix R. Project proposal exam committee .......................................................................................... 102
1. Introduction

In modern consumer electronics, conventional button-interfaces are increasingly being replaced by button-less interfaces which allow new interaction methods. Touch screens, for example, make it possible to develop interfaces which enable a direct manipulation of two-dimensional virtual objects. The development of three-dimensional tracking systems that do not require body-mounted markers has stimulated the implementation and development of gesture-based interfaces. Gesture-based interfaces can be controlled by moving a remote control (e.g. Nintendo Wii) or moving one’s own body (e.g. Microsoft Xbox Kinect) through a three dimensional interaction space. An advantage of this type of interface is that it is not bound to two-dimensional manipulation, which is expected to lead to a more natural interaction experience (Wachs, Kölsch, Stern, & Edan, 2011). In addition, it enables the user to interact with technology from a remote location without the need for additional tools, and in theory, with an infinite number of instruction possibilities. However, a disadvantage of this type of interface is that the interaction possibilities need to be explicitly taught to the users, and therefore the interaction might initially not be intuitive. Indirect manipulation of a distant object (i.e. gesture-based interaction) might be perceived as unnatural, and therefore the choice for interaction gestures needs to be made with great care. An option when investigating intuitive gestures is to look at metaphors originating from common object-use or from human language. Although common and intuitive gestures could be chosen for a specific interaction task, users should learn which intuitive gestures to use. Therefore, a learning process is still required.

This research project aims to address both issues: which gestures are intuitive and how should they be taught? First, an explorative study was performed to investigate what are potentially suitable gestures for implementation in a Philips television system. And second, it was investigated what the effects of different teaching-methods are on the user’s memory for these newly learned gestures. In existing implementations of gesture-based interfaces in games, a commonly used method is to teach users the required gesture-set by means of an ‘observation and imitation’ task, which the users perform during the introduction of the game (e.g. Microsoft Xbox Kinect, & Nintendo Wii). During such a task the user typically views a short animation of a human figure making a certain movement (e.g. moving an outstretched arm from left to right), and is subsequently asked to imitate that movement. The user does receive feedback about the performance, but no additional explicit information about the meaning of the gesture.
This teaching method does not have to be the most effective method to teach a gesture-set. Instead of showing a human arm, objects could be shown that elicit the gesture’s movement. For example, objects such as used in regular interfaces (like rotary knobs or scroll bars) could elicit pantomime movements of the object’s use. Teaching gestures in this way would take advantage of implicit knowledge of object affordances. Also, the effectiveness of various teaching methods may differ for different types of gestures, because from literature it is known that the type of gesture which is used influences the learnability and memorability of the gestures (Carmo & Rumiaty, 2009). In the current study multiple gesture teaching-methods are compared and tested, within the framework of a potential implementation of gesture-based interfaces, which is the new generation of televisions. In this context, both learning speed and memory for gestures are important variables when dealing with both first-time and experienced users. These variables could influence the attitude of the user towards a gesture-based interface, and are therefore essential for the consumer’s attitude towards the television system as a whole.

From a scientific perspective the comparison of the different teaching methods for gesture-acquisition is interesting, because it requires the combining of theories from various fields of study, from neuroscience to ecological psychology. In addition, the topic under investigation is relatively new; studies that are known to test gesture teaching-methods usually have a different focus. Most of these related studies focus on two-dimensional gestures for touch screen interaction (e.g. Bragdon, Ugray, Wigdor, Anagnostopoulos, Zeleznik, & Feman, 2010), natural occurring co-speech gestures (e.g. Beilock, & Goldin-Meadow, 2010), implications of the implementation of a gesture interface in stereoscopic contexts (e.g. Van Beurden & IJsselsteijn, 2010) or the differences in acquisition of different gesture types (e.g. Carmo & Rumiaty, 2009), but to our knowledge none of these studies investigate the effect of teaching-methods for three-dimensional interaction gestures.

The current study aims to fill this gap by looking at the two main methods to teach interaction-gestures, which are the imitation of body movements and the inference of movements based on object affordances. In the following paragraphs an overview is given of related literature and the inherent hypotheses for the current study. But it starts with an important definition of the topic under consideration: gesture interaction.
**Gestures**

The concept ‘gesture-interaction’ is interpretable in multiple ways, similar to the concept ‘gesture’. In the context of modern technologies a gesture interface is defined as an interface which allows interaction by responding to body or object movement. As mentioned before, in this study the focus is on remote body movements. More specifically, the focus is on gestures: “movements […] [without direct feedback] that express ideas or feelings” (‘gesture’; Cambridge Dictionaries Online, n.d.). But are all movements gestures? When answering this question from a solely technical perspective the answer could be yes; in theory every movement could be assigned to a certain interaction goal. However, this does not necessarily imply that the movement is meaningful to the user. Therefore, in the following paragraphs attention is paid to the use of gestures in human language.

From a linguistic point of view, gestures are an integral part of human language (McNeill, 1992), and different movements serve different communicative goals. This view is supported by the finding that mental processing of gestures seems to involve semantic processes similar to the processing of words (Willems, & Hagoort, 2007). Multiple taxonomies exist for gestures. The most basic categorization describes a continuum between spontaneous gestures and socially conventional signs. This “Kendon’s continuum”, as described by McNeill (p.37, 1992, referring to a publication by Kendon in 1988), starts with gesticulation at the spontaneous side of the continuum, and via language-like gestures, pantomimes, and emblems, ends with the most conventional gestures of sign languages. The first type, gesticulation, includes all complementing gestures which are produced simultaneously with speech. The second type, language-like gestures are gestures that can replace language structures like words, and can be produced simultaneously, or subsequent to speech (McNeill, 1992). Both these gesture types could be called co-speech gestures; gestures that accompany speech and are pragmatically, semantically, and temporally tightly integrated with speech (Özyürek, Kita, Allen, Furman, & Brown, 2005). According to more recent categorizations, four types of co-speech gestures can be distinguished: deictic gestures, iconic gestures, metaphoric gestures, and beat gestures (Cassell, McNeill, McCullough, 1998). Each type serves a different function in communication. Deictic gestures are manual pointing gestures which are intended to establish joint attention (Kelly, Barr, Breckinridge Church, Lynch, 1999). The main function of deictic gestures is to locate characters in space, and to depict the spatial relationships between them (Cassell et al., 1998). Iconic
gestures “imagically depict[s] objects, qualities, or activities” (p.579, Kelly et al., 1999) by imitating (a part) of the action. Iconic gestures can serve two functions. They can serve to elaborate on the manner in which an action is performed, or they can serve to clarify the viewpoint from which an action is described (Cassell et al., 1998). Metaphoric gestures are intended to depict a concept without a physical form (e.g. duration) (Cassell et al., 1998). Beat gestures serve to support the textual aspects of a message, but are not influenced by the content of the message (e.g. when moving hands between grammatical parts of a long spoken sentence) (Cassell et al., 1998).

The next gesture type in Kendon’s continuum is pantomime. Pantomime gestures depict actions or objects without the need for accompanying speech (McNeill, 1992). The form of the gesture is relatively unconventional, but might be affected by cultural or linguistic characteristics. In contrast, emblems (also called quotable gestures) are standardized gestures (e.g. ‘thumbs-up’) which are highly influenced by cultural conventions, and therefore have a consistent form (McNeill, 1992). This is also the case for gestures used in sign language. But in sign language each linguistic structure has its own conventional sign which makes it a complete language without any links to hearing languages (McNeill, 1992). In addition to Kendon’s continuum, one important gesture type should be added to complete the summary of gesture types. Interactive gestures, as defined by Bavelas, Chovil, Coates, and Roe (1995), are gestures which regulate dialogues (e.g. turn taking). Although humans use only few of these interaction gestures when communicating to an artificial addressee, they are being produced (Mol, Krahmer, Maes, & Swerts, 2009). In addition, it might be expected that the use of these gestures increases when humans see artificial addressees as fully-fledged interaction partners when interaction possibilities of technology improve.

Simulation theory

All gesture types consist of hand or arm movements, but they are presumed to be based on different mental processes (Bartolo, Cubelli, Della Sala, & Drei, 2003; Hostetter, & Alibali, 2008). The simulation theory sets out a structure that might be seen as the foundation of most naturally-occurring gesture types (Goldin-Meadow, & Beilock, 2010, also Wagner-Cook, & Tanenhaus, 2009, Hostetter, & Alibali, 2008). Simulation, as described in the simulation theory, is a derivative of the grounded cognition approach to studying cognition, and is increasingly seen
as one of the main processes by which the brain handles the information coming from the world outside (Barsalou, 2009). The process of simulation is defined as “the re-enactment of perceptual, motor and introspective states acquired during experience with the world, body and mind” (p.618, Barsalou, 2008). This implies that an object or situation has been experienced beforehand, and the resulting sensations have been stored in memory. All instances of an experience with a certain object or situation are integrated by a process called the simulator, to form a distributed multi-modal system (Barsalou, 2009). This system is used to generate (a subset of) the simulation when activated. However, this activation does not result in a complete reinstatement of the modal states. To give an example: the ‘pencil’ concept would be formed by all previous experiences with the tool, and contains modality specific information like visual aspects (e.g. shape), sounds, smell, touch, taste, emotional associations and other internal states, and potential actions which it can serve (including the inherent motor states). This simulation is said to be conceptualized in a situated manner (Barsalou, 2009), which means that it takes place in a certain context, similar to activities in the outside world. When looking at the ‘pencil’ example, the context would be the situation in which the pencil could be used, and this situation is viewed as an inherent part of the simulation. So, this simulation might simulate the experience of writing a specific word on a piece of paper in a faded manner.

It is suggested that the activation of such a simulation can prime actions or inherent associations (Hostetter, & Alibali, 2008). This priming process might explain the strong tendency of humans to imitate actions performed by others (Tessari, Canessa, Ukmar, & Rumiati, 2007). In addition, it is the basis for the view on the origin of gestures stating that “gestures are an outgrowth of simulated action and perception” (p.668, Goldin-Meadow, & Beilock, 2010). A good example of this prompting process can be seen in the study by Wagner-Cook and Tanenhaus (2009) in which participants were asked to solve the Tower of Hanoi problem. The participants had to move discs of different sizes from and to three different pegs. The problem was solved in one of two possible conditions. In the first condition, the participant had to pick up real discs and move them from one peg to the other. In the second condition, the participants solved the problem by dragging virtual disks horizontally from one position to the other. In this condition they used a mouse to interact with the computer. After the task the participants were asked to explain how they solved the problem. The results showed that the condition a participant was in did not have an effect on the verbal descriptions given, nor on the number of
gestures used. However, there was an effect on the gestures they used. Participants in the ‘real disk’ condition used more grabbing movements and curved trajectories in their gestures. On the other hand, participants in the ‘virtual disk’ condition used more horizontal and linear movements (Wagner-Cook, & Tanenhaus, 2009). These results were interpreted as being the reflections of kinematic characteristics of the motor plan that was used to solve the problem (Wagner-Cook, & Tanenhaus, 2009).

According to the simulation theory, observing direct manipulation of an object by another person results in a direct mental simulation of the action (Barsalou, 2009). The discovery of mirror neurons in cortices such as the primary motor cortex of macaque monkeys has been put forward as support for this theory (Cartmill, Beilock, & Goldin-Meadow, 2012; Barsalou, 2008). These mirror neurons are neurons that fire when the monkey performs an action and when it observes the same action being performed by someone else (Hickok, 2008). They are expected to exist in the human brain as well. Mirror neurons in a primate do not fire when the action is being pantomimed, or when the monkey observes only the targeted object (Hickok, 2008). However, some preliminary results suggest that there are certain types of mirroring processes within the human brain that do respond to pantomime and objects (Cartmill et al., 2012).

To conclude, the exact role of mirror neurons in simulation processes is still a controversial topic (Hickok, 2008). In contrast to what would be assumed when looking only at the studies on mirror neurons of monkeys, simulations in humans are, by some views on simulation processes, also assumed to be activated when observing objects that afford activities, and when observing pantomimed or intransitive (non object-based) movements (Barsalou, 2009). This would imply that all teaching-methods tested in the current study would result in mental simulations of gestures. When observations of direct manipulation, pantomimes, objects, and intransitive movements all result in mental simulation, the results of the current study probably cannot distinguish the kind of process underlying the learning of gestures.

**Observational learning**

In this study two types of observational learning are tested; learning by imitation and learning by object observation. The process of imitation is being described by the dominant view of the dual-route model of imitation. According to this model, the imitation of meaningful actions is facilitated by the retrieval of these actions from memory (Carmo, & Rumiati, 2009). The
imitation of meaningless actions can not be facilitated by stored actions, but uses more complex visuo-motor integration processes. This division could explain why the imitation of meaningful actions is more accurate than the imitation for meaningless actions when cognitive resources are low (Carmo, & Rumiati, 2009). On the other hand, when meaningless actions are learned and are stored in the long-term memory, performance is more accurate than for meaningful actions (Tessari, Bosanac, & Rumiati, 2006). This finding was explained by the authors by the limited number of instances that led to the memory of meaningless actions, relative to the meaningful actions of which the memory could have been influenced by numerous action, or gesture observations (Wagner-Cook, & Tanenhaus, 2009). The implication of these findings might be that when for a certain interface a precise gesture form is required it might be better to choose meaningless gestures. However, these gestures will require a longer learning process.

This trade-off between a precise and (computer) interpretable gesture form and the intuitiveness of a gesture, indicates the focus of today’s studies focusing on the development of gestures-based interfaces. New techniques are being developed to enable accurate and robust sensing of hand gestures (for an overview of techniques see Wachs et al., 2011; Cohn, Morris, Patel, & Tan, 2012). In addition, to our knowledge studies on the implementation of intuitive gestures are still mainly performed for touch-screen applications (e.g. Wobbrock, Ringel Morris, & Wilson, 2009), or for object-movement-based gestures (e.g. Ruiz, Li, & Lank, 2011). Both studies by Wobbrock and colleagues (2009) and Ruiz and colleagues (2011) deployed a method in which participants improvised gestures for specific tasks. In the study by Wobbrock et al. (2009) participants were asked to manipulate a number of objects on a touch-screen. For part of the manipulations strong agreement existed between participants. Interestingly, these agreements were based on the Windows desktop metaphor for almost three out of four gestures. The authors concluded that only a subset of the tasks, for which there was a general agreement between participants, was suitable for an intuitive gesture-based interface. In the study by Ruiz et al. (2011) participants were instructed to improvise and make arm movements for smartphone control tasks while holding a smartphone. The results showed large agreement between participants for common control-tasks. In addition, these gestures were mainly pantomimes of conventional phone interactions, or were based on the physical object-manipulation metaphor (Ruiz et al., 2011). An attempt to implement intuitive remote gestures by using available gesture sensing-technologies is the study by Song, Goh, Hutama, Fu, and Liu (2012). In their study the
skewer metaphor was used to enable participants to manipulate various virtual objects by using both hands. This method is feasible from a technological perspective, because currently available vision-based gesture sensing technologies can robustly sense hand position, while hand orientation is still problematic. Although the metaphor of a skewer was not based on a user-centered study it was relatively easy for the participants to understand. However, it should be mentioned that the participants of this study received direct visual feedback on their manipulations, which makes the gestures intrinsically different from the gestures under consideration in the current study.

When looking at the process of imitation, an additional separation should be made between the imitation of transitive (object-related) actions and intransitive (not object-related) actions. Imitation of transitive actions is less accurate than the imitation of intransitive actions. This could be explained by the intrinsic complexity of these movements (Carmo, & Rumiati, 2009). In addition, this could also be explained by the distinction between ‘means’ and ‘end’ processes during action observation. It is suggested that when observing an action the human mind is mainly interested in the action’s goal (‘end’) and subsequently in the means (De Lange, Spronk, Willems, Toni, & Bekkering, 2008; Bekkering, Wohlschläger, & Gattis, 2000). The goal of an action can already be known by the observer or could be inferred during action observation. In a study by Cuijpers, Van Schie, Koppen, and Erlagen (2006) a computational model is presented describing how the goal of an action might be inferred from action observation. In this model the observer’s task knowledge, the environment and the observer’s action repertoire are used to judge the likelihood of certain action goals when observing the movement kinematics (Cuijpers et al., 2006). When solely observing a hand movement, as is the case for imitation of intransitive movements, the environment component of the model cannot be used to infer the goal. This would imply that the processes rely on the observer’s task knowledge and action repertoire. In addition, when the task knowledge is insufficient, as might be the case for unintuitive or meaningless gestures, inferential processes would rely even more on one factor, which is expected to reduce the efficiency of the inference processes significantly. Therefore, only the observation of hand movements depicting intuitive gestures is expected to combine the factors as described by the model, whereas for unintuitive or meaningless actions this is not expected. This reasoning is in line with the dual-route model of imitation (Carmo, & Rumiati, 2009).
The other method under consideration is ‘learning by object observation’. An assumed advantage of this method is that the actions goal is clearly visible. In addition, the object’s action possibilities can be communicated through its affordances; “the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used” (p.9, Norman, 1988). These affordances play an essential role in the everyday interaction with objects. Affordances are being used in design processes of both physical and virtual products. However, currently there are no instances known that make use of object affordances to elicit remote gestures. A remarkable example of affordance-use for multi-touch gestures is ‘Affordance Play’ (Bragdon et al., 2010) in which basic touch-gestures are being taught to the user by means of an introduction game. In this game the directions of swipe-movements were being afforded by showing affordable objects (e.g. springs and pads). The results showed that such an interface resulted in more practicing behavior but not in an improved recall (Bragdon et al., 2010). A main problem with this approach is that the affordable objects that were used were not related to the task itself. This caused the objects to be unintuitive, and therefore it did not aid the memory for the specific tasks.

The gestures under consideration in the current study are remote standalone gestures without direct feedback. When the object-observation method is used, the resulting gestures are pantomimes. Pantomimes are gestures which are formed in the working memory and involve stored information of perceptual object characteristics, action semantics and stored procedural programs (Bartolo et al., 2003). Pantomimes are a special type of gestures for which categorization is relatively complex. For example, the transitiveness of pantomimes is debatable; pantomimes refer to an object, but the object is not physically present. This latter characteristic causes problems to the use of affordance theories, when certain definitions of object affordances are being used. Earlier theories on affordances limited the effect of object characteristics only to visible objects (e.g. Gaver, 1991), but more recent studies have found that also references to objects can trigger the objects’ affordances (Barsalou, 2009). In a study by Glover and colleagues (Glover, Rosenbaum, Graham, 2004), participants were instructed to read a word and subsequently grasp a wooden block. When the word referred to a small object (e.g. grape), the size of the grip aperture was smaller at the onset of the grasping task than when the word referred to a large object (e.g. apple). These micro-affordances of ‘invisible’ objects demonstrate the important role of association on the activation of action components (Ellis, & Tucker, 2000).
These findings led to two main expectations related to the results of the current study. Firstly, because pantomimes are intrinsically associated to objects, the affordances of these objects are expected to lead to automatic processing, similar to the effect of the affordances of physical objects. In addition, it is expected that these affordances will result in clear communication of the required movement (i.e. the gesture to be performed). Secondly, because it is assumed that the human mind is at first interested in the goal of a movement (De Lange et al., 2008; Bekkering et al., 2000), and the pantomime method communicates a clear goal, it is expected that the pantomime teaching-method results in better gesture memory than the imitation teaching-method.

When imitation and pantomime tasks are combined (in a situation were both hand and object are visible) the effect of object affordances is found to be relatively large. In a study by Bach, Bayliss, and Tipper (2011) affordances of task irrelevant objects significantly modulated the estimation of grip-aperture size during imitation. This is in line with the view that human imitation processes are intrinsically goal directed (Bekkering et al., 2000), even if a person does not consciously attend to the goal (De Lange et al., 2008). These processes are expected to serve motion anticipation similar to what is assumed for simulations and forward models (Barsalou, 2009) during action observation. This influential power of object affordances resulted in the expectation that a combination of the imitation and pantomime condition would not lead to an improved recall compared to the pantomime condition.

**Practice**

It is commonly accepted that ‘practice makes perfect’. This saying is even applicable to the activation of mirror areas by practiced and unpracticed novel (dance) movements (Cross, Hamilton, & Grafton, 2006; Thornton, & Knoblich, 2005). Observation of rehearsed or mastered movements resulted in greater activity in the resonance areas compared to observation of unrehearsed movements. In addition to these examples, the ‘enactment paradigm’ describes the effect of physical practice on memory (Tessari et al., 2006). Memory is better for tasks that are physically performed (e.g. ‘open door’), than for observations, or descriptions of these tasks (Tessari et al., 2006). A study by Wagner Cook, Yip, and Goldin-Meadow (2010), revealed that this “doing component” (p.465) can also be added to memory tasks in other modalities. Participants who described a list of images by using gestures (spontaneous or forced) recalled more images in a subsequent recall task than participants who did not gesture. However, the
opposite is true as well; not practicing can improve memory as well (Cohen, Pascual-Leonie, Press, & Robertson, 2005). In a study on the effect of consolidation after practice it was found that procedural motor information could improve over the day, while information related to a movement’s goal could enhance over a night of sleep (Cohen et al., 2005).

The current study consists of two experiments. The first experiment explored which gestures might be suitable for controlling a television with a gesture-interface. For this experiment gestures were defined as all movements of the body intended to control the television. The results of this experiment could be used during the design of a gesture-based interface, but they were also essential for the second experiment, because they provided a realistic gesture-set. The second experiment tested which type of gesture teaching-method resulted in the best learnability and memorability of a gesture-set. Learnability was defined for this experiment as the average percentage of correct gesture recalls in subsequent recall tasks. Memorability was defined here as the percentage of correct recalls after the distraction task, as specified in the ‘method’ section. Based on the results of the first experiment, a distinction was made between different gesture types when comparing learnability and memorability. In addition, two independent variables were manipulated; gestures were either intuitive or unintuitive, and participants were instructed to practice or not to practice.

A total of five main research questions were addressed; two for the first experiment and three for the second experiment.

Research questions experiment 1:
1. What are the commonalities of intuitive gestures for television control gestures?
2. What types of associations underlie intuitive gestures?

Research questions experiment 2:
1. Which method, the observation of a human movement or an object, supports learnability and memorability of interaction gestures most?
2. What is the effect of intuitiveness of gestures on the effectiveness of the teaching methods in terms of both learnability and memorability?
3. What is the effect of practice on the memory of gesture-sets taught by both methods?
Three hypotheses were formulated for the second experiment. First, it was hypothesized that observation of objects (pantomime method) results in better recall for gestures (i.e. both learnability and memorability) compared to the observation of human arms (imitation method). A combination of both methods is expected to result in a similar performance as the pantomime method. In addition, it was hypothesized that the pantomime method results in recall for emblems similar to recall in the imitation condition. However, a higher recall percentage is expected for object-based gestures in the pantomime condition compared to the imitation condition. Second, it was hypothesized that meaningful gestures are better recalled in all conditions, and have a shorter response time at recall. This was expected for both gesture types. And third, it was hypothesized that practice has a positive effect on the recall and response time for gestures learned in the imitation condition. The effect of practice is expected to result in less improvement for the pantomime condition than for the imitation condition.

In the first study potential gestures were explored by instructing fifteen participants to think of suitable gestures for fourteen television control-tasks which they had to perform for four different scenarios (called use-cases). During a subsequent interview the participants were asked for their inherent motivations and associations. The resulting video data were transcribed into a dataset including the body-part movements and positions included in the improvised gestures, and the inherent associations. From this dataset, frequency tables were created showing the improvised gestures per task, and a general overview was made of the used association types.

The hypotheses for the second experiment were tested in an experiment in which a predefined gesture-set, consisting of ten gestures, was being taught to 120 participants by means of one of three teaching methods. In the ‘imitation’ method participants observed human arm movements, in the ‘pantomime’ method participants observed animated interface objects, and in the combination method (‘both’) participants simultaneously observed imitation and pantomime stimuli. Gesture acquisition was tested by subsequent recall tasks in which both correct recall and response time were measured.
2. Experiment 1

2.1 Method

Design

During the experiment each participant performed 56 tasks after hearing four different scenarios (‘use-cases’) in a within-subjects design. The following tasks were performed: activate the gesture interface (1), de-activate the gesture interface (2), switch on (3), switch off (4), switch to the previous channel (5), switch to the next channel (6), mute (7), un-mute (8), turn the volume down a little (9), turn the volume down much (10), turn the volume up a little (11), turn the volume up much (12), select the video in the bottom row second from the right (13), and open the selected video file (14). However, during the trials of four out of fifteen participants, task number 13 and 14 could not be performed, because of technical problems with NetTV. Because of this, these trials could not be incorporated in the results.

The participants performed the tasks after hearing four scenario descriptions. The complete description of the scenarios ‘working’, ‘watching a movie’, ‘tidying the living room’, and ‘walking to the door’ can be found in appendix A. During each scenario condition, participants were positioned in different locations in the lab (see appendix D). During the ‘working’ condition participants were sitting at a table with a laptop. During the ‘watching a movie’ condition participants sat comfortably on the couch. And, during the ‘tidying the living room’ condition participants stood in front of the television, at a distance of one meter, and were carrying a box. The box was used to restrict the movement possibilities of the participants and force them to use only one hand. During the ‘walking to the door’ condition participants stood further from the television, next to the door.

The participants were exposed to the scenarios in random order, as well as to the fourteen tasks within each scenario.

Participants

Fifteen participants took part in this study; five females (average age 27 years old, ranging from 22 to 43 years old), and ten males (average age 25, ranging from 20 to 48). Thirteen participants were of Dutch nationality and two of Indian nationality. All participants had normal or corrected
to normal vision, no physical handicap, and were all registered in the JF Schouten School database of participants. Students of the Eindhoven University of Technology received an incentive of ten Euros for their participation; other participants received twelve Euros.

**Apparatus & Set-up**

Participants were asked to stand or sit at different locations in a living room environment, dependent on the condition they were in (for an overview see appendix D). In the room was a Philips television which was turned on during the first part of the experiment in which the participants performed the tasks. The television displayed either national broadcasts or the YouTube homepage by using the Philips NetTV connection (the latter was displayed only during the ‘video selection’ tasks).

During the experiment two cameras were used, one to record the movements participants made (A) and one to record the subsequent interview (B, see appendix D). After the interview the participants were asked to fill in a questionnaire about general demographics and previous experience with gesture based interfaces (see appendix B).

**Procedure**

Participants were given a short description about the general procedure and subsequently were asked to fill in an informed consent form (appendix C). Before the main experiment started the procedure was verbally explained in more detail and questions could be asked. In the first part of the experiment participants were asked to stand or sit in a predefined position in the living room, as determined by the scenario. The scenario (see appendix A) was read out loud by the experimenter, and the participants were asked to keep the scenario in mind while performing the subsequent tasks. Each description was followed by fourteen tasks in which the participants had to think of and perform a gesture they thought was suitable for the specific task and scenario.

After the participants performed all tasks they were being questioned about their reasons for performing these gestures, and their inherent associations. Subsequently the participants were asked to fill in the questionnaire. When the participants had finished they were thanked and paid for their participation.
2.2 Results

In this chapter the results are described per task. The gestures that were made for each task have been categorized, based on the body parts involved in the movement, the general direction, the speed, and the association the participants reported with the movement. For each task the frequency tables are provided, and the three most common gestures have been summarized and are, when necessary, illustrated by screenshots.

Commonalities

Task 1: ‘Activating the gesture-interface’

The frequencies of gestures and the distribution over participants are shown in Figure 1. The most commonly used gestures to activate the gesture-interface were:

1. To wave hands,
2. To snap fingers, and
3. To clap hands

The variation between participants for these three gestures was relatively low. In addition, participants preferred to use a single gesture for both the activation and de-activation of the gesture-based interface.

Figure 1. Distribution of improvised gestures for the 'activation' task. White lines in bars indicate contribution of different participants. N =15
**Task 2: ‘De-activation of the gesture-interface’**

The frequencies of gestures and the distribution over participants are shown in Figure 4. The most common used gestures to de-activate the gesture interface were:

1. To form or draw a cross with hands or fingers (Figure 2) (one out of three participants)
2. To clap hands (four out of fifteen participants performed this gesture)
   - To wave hands, snap fingers or to make a movement downwards (for latter, see Figure 3)

---

**Figure 2. Example of gesture category 'cross'**

**Figure 3. Example of gesture category 'make a movement downwards'**

**Figure 4. Distribution of improvised gestures for the ‘de-activation’ task. White lines in bars indicate contribution of different participants. N =15**
Task 3: ‘Switching on the TV’

The frequencies of gestures and the distribution over participants are shown in Figure 6. The most common gestures used to switch on the TV were:

1. To snap fingers
2. To open a hand (see Figure 5)

The variability for gestures made in the ‘switch on’ task is relatively low. The top two of most common gestures were performed by respectively five or four out of fifteen participants.

Figure 5. Example of gesture category ‘opening hand’

Figure 6. Distribution of improvised gestures for the ‘switch on’ task. White lines in bars indicate contribution of different participants. N =15
Task 4: ‘Switching off the TV’

The frequencies of gestures and the distribution over participants are shown in Figure 9. The view of the participants on the relation between the gesture for switching on and switching off was very divers, no clear preference was found. The most commonly used gestures to switch off the TV were:

1. To move one or both arms horizontally (see figure 7)
2. To show an open hand, palm directed towards the TV (see figure 8), or
3. To snap fingers

Figure 7. Example of gesture category ‘horizontal arm movement’  Figure 8. Example of gesture category ‘show hand’

![Figure 7](image1.png) ![Figure 8](image2.png)

Figure 9. Distribution of improvised gestures for the 'switch off' task. White lines in bars indicate contribution of different participants. N =15
Task 5 & 6: ‘Going to the previous channel’ & ‘Going to the next channel’

The frequencies of gestures and the distribution over participants are shown in Figure 10 (‘previous channel at the left, ‘next channel’ at the right). Similar to the most common gestures for the task ‘next channel’ the most common used gestures to switch to the previous channel were:

1. To swipe (dragging a horizontal list of channels forward or backwards), and
2. To point to the next channel (to the left for previous or to the right for next)

The variability of the improvised gestures for this task was very low. Nine out of fifteen participants used a ‘swiping’ gesture to change channels, and six out of fifteen a pointing gesture. The relation between the gesture for instructing to switch to the next channel or to the previous channel was in all instances opposite.

![Figure 10. Distribution of improvised gestures for the ‘previous channel’, and ‘next channel’ task. White lines in bars indicate contribution of different participants. N =15](image)
Task 7: ‘Muting the TV’

The frequencies of gestures and the distribution over participants are shown in Figure 12. The most common gestures used to mute the television were:

1. To show a closing a hand (see Figure 11), and
2. To block an ear with a hand (see Figure 13 for the opposite movement)

The variability for both the mute and un-mute task was relatively high. Participants mentioned that it was difficult to think of suitable gestures for these tasks. However, most participants agreed that the gestures for mute and un-mute should be opposite to each other.

Figure 11. Example of gesture category 'closing hand'

Figure 12. Distribution of improvised gestures for the 'mute TV' task. White lines in bars indicate contribution of different participants. N =15
**Task 8: ‘Un-muting the TV’**

The frequencies of gestures and the distribution over participants are shown in Figure 14. The most common gestures used to un-mute the television were:

1. To show an opening a hand (see Figure 11 for the opposite movement), and
2. To show a movement aimed to unblock an ear with a hand (see Figure 13)

![Un-mute TV](image)

*Figure 13. Example of gesture category ‘unblocking ear’*

![Un-mute TV](image)

*Figure 14. Distribution of improvised gestures for the ‘un-mute TV’ task. White lines in bars indicate contribution of different participants. N =15*
Task 9: ‘Decreasing the volume a little’

The frequencies of gestures and the distribution over participants are shown in Figure 15. The most common used gestures to decrease the volume a little were:

1. To move an arm vertically (downwards)
2. To pantomime the movement of turning a rotary knob (see Figure 16), and
3. To show a vertical oriented size by showing a partly opened hand (see Figure 17)

The variability was relatively low. In addition, most of the participants agreed that the relation between the gestures for increasing and decreasing the volume was opposite. The amount of volume change could be indicated by the size, speed, or by a repetition of the gesture.

Figure 15. Distribution of improvised gestures for the ‘volume down a little’ task. White lines in bars indicate contribution of different participants. N =15
Figure 16. Example of gesture category ‘turning a rotary knob’

Figure 17. Example of gesture category ‘partly opened hand’

Figure 18. Example of gesture category ‘vertical arm movement’
Task 10: ‘Decreasing the volume much’

The frequencies of gestures and the distribution over participants are shown in Figure 19. The most common used gestures to decrease the volume much were:

1. To move an arm vertically (downwards) (see Figure 18 for opposite movement), and
2. To pantomime the movement of turning a rotary knob (see Figure 16)

Task 11: ‘Increasing the volume a little’

The frequencies of gestures and the distribution over participants are shown in Figure 20. The most common used gestures to increase the volume a little were:

1. To move an arm vertically (upwards) (see Figure 18)
2. To pantomime the movement of turning a rotary knob (see Figure 16 for opposite), and
3. To show a vertical oriented size by showing a partly opened hand (see Figure 17)

Task 12: ‘Increasing the volume much’

The frequencies of gestures and the distribution over participants are shown in Figure 21. The most common used gestures to increase the volume much were:

1. To move an arm vertically (upwards) (see Figure 18)
2. To pantomime the movement of turning a rotary knob (see Figure 16 for opposite), and
3. To point up
Figure 19. Distribution of improvised gestures for the 'volume down much' task. White lines in bars indicate contribution of different participants. N = 15

Figure 20. Distribution of improvised gestures for the 'volume up a little' task. White lines in bars indicate contribution of different participants. N = 15

Figure 21. Distribution of improvised gestures for the 'volume up much' task. White lines in bars indicate contribution of different participants. N = 15
**Task 13: ‘Selecting a video file’**

The frequencies of gestures and the distribution over participants are shown in Figure 22. The most common gestures used to select a video from the grid of videos were:

1. To point to the video, and
2. To make horizontal and vertical curved linear movements to illustrate the steps required to move a ‘current selection’ to the dedicated file (“jumping highlight”) (see Figure 22)

The variability for this task is very low; most participants choose either pointing or stepwise instruction, or used a combination of both across the four conditions.

![Figure 22. Example of steps involved in 'jumping highlight' gesture (blue line) in which the participant points and jumps to the individual icons on screen. Jumps can be both directed towards the z-axis or can be drawn on the plane of the television screen.](image)

![Figure 23. Distribution of improvised gestures for the 'select a video file' task. White lines in bars indicate contribution of different participants, N =11](image)
Task 14: ‘Opening a video’

The frequencies of gestures and the distribution over participants are shown in Figure 24. The most common gestures used to open the selected video were:

1. ‘To press’ or to make a short movement directed towards the television
2. To keep the arm a few seconds in the current position (‘no movement’), and
3. To increase the size between the fingers (“similar to the zoom-gesture used for touch-screen interfaces”, participant P8) (see Figure 23)

Figure 24. Example of gesture category 'increasing size'

Figure 25. Distribution of improvised gestures for the 'open a video file' task. White lines in bars indicate contribution of different participants. N = 11
Association types

Participants reported different types of associations for the various tasks. In this experiment a distinction was made between four types of associations. The first type ‘interface object’ included all gestures that participants associated with the interaction with objects or interfaces (e.g. the pantomime of rotating a rotary knob). The second type ‘common gestures’ included all gestures that were standard in human language (e.g. thumbs-up). ‘Abstract conceptual model’ gestures were based on associations that were not related to the current task, other object, or interface, nor were common gestures (e.g. drawing a semi-circle in reference to the sun-set). The fourth type, ‘direction’, includes all pointing gestures. An additional fifth type was created which included all gestures for which participants could not give an association. For each task-group the percentage of reported association-types is shown in Figure 25. As can be seen in Table 1, most gestures were based on associations with interface objects. The percentage of interface associations was related to the perceived difficulty of each task group. Tasks that were perceived as being easy resulted in more ‘interface object’ gestures. More difficult tasks resulted in more ‘common gestures’ and in gestures without a clear association. Participants reported the volume control and switching channels tasks as being relatively easy, while they reported that muting and activation tasks were relatively difficult.

![Distribution association type per task](image)

Figure 26. Distribution of used gesture association-types per task-group. For an elaborate explanation of the association types see caption of Table 1.
Table 1. Overall frequency of association-types. Total number is less than total number of gestures made, because participants did not report an association for all gestures made. Reported association types: ‘interface object’: association with other product or interface, ‘common gesture’: association with and use of cultural standard gesture, ‘abstract conceptual model’: association with situation or object not related to current task, other product or interface, ‘direction’: gesture indicating direction and is not associated with a product or interface.

<table>
<thead>
<tr>
<th>Association Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface object</td>
<td>247</td>
</tr>
<tr>
<td>Common gesture</td>
<td>157</td>
</tr>
<tr>
<td>Don’t know</td>
<td>73</td>
</tr>
<tr>
<td>Abstract conceptual model</td>
<td>50</td>
</tr>
<tr>
<td>Direction</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>530</strong></td>
</tr>
</tbody>
</table>

Scenario differences

Figure 26 shows the differences between the gesture-category distributions for the four scenarios. There was relatively little variation between the gesture distributions for the scenarios. The ‘tidying the living room’ caused most variation, because participants could only use one hand in that condition, whereas there were no restrictions in the other scenarios. Gestures that were performed with two hands in the other conditions were transformed into a one-hand gesture or abandoned in the ‘tidying the living room’ condition.
Figure 27. Distribution of used gestures per context.
2.3 Discussion

The goal of this preliminary test was to explore which gestures might be suitable to control a television with a gesture-based interface. Three variables were used to determine suitability. The first variable was ‘frequency of occurrence’, because this could give an indication of general association. The second variable was ‘association type’, because this variable could be used to find potential type preferences or task correlations. Finally, to find potential restrictions or preferences for certain contexts, ‘scenario’ was added as the third variable.

An experiment was conducted in which fifteen participants made gestures for fourteen tasks, after hearing one of four scenario descriptions. The results show clear commonalities between the gestures produced by different participants. These patterns are best visible in the ‘switching channels’ and ‘volume control’ task-groups, which participants reported to perceive as being relatively easy to perform.

The perceived difficulty of the different tasks was related to the amount of similar gestures that were made, as well as to the types of gestures that were made (also see Figure 26). Tasks that participants perceived as being relatively easy were more often associated to other products or interfaces, which resulted in pantomime gestures referring to object-use. This might suggest that the use of a suitable interface metaphor is preferred for basic TV-instruction tasks. Tasks that were perceived as being more difficult, like ‘muting’ and ‘switching on/off’ tasks, were more often paired with a gesture that was common in human interaction, especially emblem-like gestures. Other types of gestures were hardly used by the participants to control the television, which suggests that further investigation should focus on the implementation of pantomime gestures and emblems. A summary of the gestures that were most often used per task can be found in Appendix F.

The gestures did not show large differences for the various scenario conditions. Most participants preferred to make similar gestures for each task in all scenario conditions. Only for participants who chose two-handed gestures the condition in which they could only use one hand was perceived as problematic, and resulted in adjusted or other gestures. These results suggest that context difference probably will not affect the suitability of gestures which are performed with one hand.
3. Experiment 2

The main goal of this second experiment was to test which teaching-method is most effective for teaching a gesture-set. This was done by teaching a set of ten gestures to each participant. The gestures that were used were intuitive gestures that mainly originated from the results of the first experiment. In addition, based on the results of the first experiment, it was decided to select both pantomime and emblem gestures to be able to control for the potential effect of gesture type. However, not all gestures could be used, because the ‘pantomime’ teaching-method of the current experiment required the gesture to be transformed in an object that affords the movement inherent to this gesture. Because there were not sufficient useable gestures, two new gestures were added. These two emblem gestures (‘stop’ and ‘number two’) have been selected because of their widespread meaning, and their results in two preliminary tests (see appendix G).

3.1 Method

Design

The experiment had a 3 teaching-method (imitation/pantomime/both) x 2 intuitiveness (intuitive/not intuitive) x 2 practice (practice/no practice) between subjects design. The conditions of the independent variable teaching-method are illustrated by Table 2. The independent variable intuitiveness determines the task association per gesture; gestures can either have an intuitive or an unintuitive task association. The intuitive task associations were based on the results of the first experiment. The unintuitive task associations were associations with

<table>
<thead>
<tr>
<th>imitation</th>
<th>pantomime</th>
<th>both</th>
</tr>
</thead>
</table>

Table 2. Overview of conditions inherent to the independent variable teaching-method. The table shows screenshots of the stimuli for the gesture for 'two'.

32
gestures from the set which normally would not be associated to the current gesture; task associations of continuous tasks were paired with gestures indicating dichotomous tasks and visa versa. For an overview of both intuitive an unintuitive task associations per gesture see appendix J.

There were two dependent variables: the percentage of correctly recalled gestures per test and the response time per gesture. A gesture could only be judged as being completely correct or incorrect. This judgment was made based on a list of criteria that is presented in appendix K. The response time was defined as the time between the presentation of the task and the onset of the correct gesture in milliseconds.

During the experiment each participant was presented with a gesture-set consisting of ten items. An overview of the gesture stimuli can be found in appendix J. The choice of items was based on experiment 1 and on two preliminary tests. The first experiment was conducted to investigate which gestures were associated with a number of television control-tasks. In addition to potential stimuli, this test revealed a potential confounding variable for the current experiment in the type of gesture. Therefore, it was chosen to select five object-based gestures and five emblems, to be able to control for gesture-type. The second and third preliminary tests were aimed to test and iterate custom-made stimuli on interpretability, perceived difficulty, and potential overlapping associations. For a complete description of preliminary tests two and three see appendix G.

Participants
120 participants took part in the experiment; 45 females and 75 males. The average age of female participants was 23 years with a standard deviation of 5 years. The average age of male participants was 24 years and also had a standard deviation of 5 years. All participants were selected to be right handed (based on self report) and had normal or corrected to normal vision. The participants received an incentive of five (for university students) or seven Euros for their participation.

Set-up & apparatus
The experiments were conducted at the PsyLabs of the University of Technology Eindhoven. Four participants could participate simultaneously. During the experiment the participants were
standing in a room in front of a desk with a keyboard, a monitor (19 inch, resolution 1280x1024), and a form for the distraction task. A webcam (recording resolution 640x480) was mounted on top of the monitor, and registered the arm movements of the participants. For a schematic overview of the set-up see appendix M.

A custom-made SWF program was built (by using Adobe Flash CS5 software) to instruct the participants, present the stimuli, (video) record the responses, and measure the response times.

As can be seen in Table 2, three types of stimuli were used during the experiment. The stimuli for the imitation condition were video taped and edited to match the stimuli of the pantomime condition (by using Adobe Premiere CS5). The stimuli for the pantomime condition were digitally animated by using Adobe After Effects CS5 software. The durations of the animated stimuli were determined by investigation of the inherent video taped stimuli. For an overview of the stimuli characteristics of the individual items see appendices J and L.

Participants were asked to fill in a questionnaire, in which they were asked for their general demographics, for an additional self report for handedness, and for a potential sign-language skill. This questionnaire was filled in before the start of the distraction task. The distraction task was a memory task in which participants had to recall briefly presented sequences of figures. In total seventeen sequences of different length had to be recalled. For the form containing the questionnaire and the distraction task see Appendix N.

Procedure

Participants were welcomed and received a general introduction to the procedure of the experiment. After this they were asked for their informed consent. Participants were randomly assigned to a condition and room and received additional instructions from the experimenter. The participants were not aware of the conditions in the experiment, and neither did they know the conditions they had been assigned to. They were instructed to stand at a marked location in the room and were explicitly told by the experimenter only to use their right arm during the experiment. After this, the participants continued individually.

The experiment started with an elaborate instruction in which the participants were told either to practice or not to practice (depending on the condition) while viewing the stimulus videos (for an overview of the instructions per condition see Appendix P). When the participants indicated that they were ready, the first part of the experiment started, and participants viewed
the ten videos in random order. After viewing the videos the participants were instructed to hold the spacebar of the keyboard with their right hand until they saw the name of the gesture they had to make on screen and knew which gesture to make. Then, participants released the spacebar and made the gesture with their right hand before pressing and holding the spacebar again. This procedure was repeated for all ten gestures in random order. The complete cycle of viewing and recalling the items was repeated twice; participants observed the sequences and performed the recall task three times during this part of the experiment (i.e. test 1, test 2, & test 3). After the third test the instruction appeared on the screen to start with the questionnaire and subsequent distraction task. Participants were not aware that a subsequent gesture recall task would follow the distraction task.

The questionnaire and distraction part of the procedure was performed on a printed form, and participants were instructed to sit down while performing the tasks. During the distraction task participants viewed a sequence of figures (squares, circles, diamonds, and crosses) on screen for a short period of time, and were instructed to draw as many figures on the sheet of paper as they could remember. The distraction task started easy with sequences consisting of only two figures, but rapidly increased to eight figures per sequence. After viewing all eighteen sequences a message appeared on screen that they had to stand at the marked positions to perform a final gesture recall task (test 4). After this fourth test the participants were thanked for their participation and received their incentive. The duration of the whole experiment was approximately thirty minutes.

Annotation & data analysis
The analysis of the data was preceded by an extensive process in which the conditions were verified and the recalled gestures were judged after observing the 5160 video files. A short description of each step in this process is given below.

*Stimulus presentation observation* - A video recording was made each time the participants viewed the ten stimulus videos to check if participants did or did not practice, as was instructed at the beginning of the experiment. Participants who did not act during the first viewing as was instructed were reassigned to a condition to match their behavior. Three participants could not be
assigned to a condition, because their hands were not visible on the recording, and were therefore excluded.

If participants performed practicing behavior during the second and third viewings inconsistent to previous practicing behavior, these subsequent blocks were discarded. This was the case for nineteen participants during the second test, and six participants during the third test.

*Recall observation* – The gestures which the participants recalled were automatically recorded when the participants released the spacebar. However, ten participants misunderstood the instructions which caused that their gestures were not recorded and therefore could not be analyzed. All other recordings were viewed and according to the predefined criteria the arm movements were judged as being correct (‘1’) or incorrect (‘0’). The criteria that were used for these decisions can be found in appendix K. Three items were found to be frequently interpreted in an unintended manner in the pantomime condition (gesture 1, 2, and 7). This could be due to the limited amount of iterations in the preliminary tests. To be able to judge the memory for these gestures as well, the criteria were slightly adjusted (see Appendix K for the original and adjusted criteria).

*Response time* – The response time data for all correctly recalled gestures were used for the analysis. Response time data was deleted when a recall was incorrect, when the response time was too short to be a valid data point (less than 150 ms, the minimal required time to respond to a visual stimulus), or when the value was extremely large (more than three standard deviations from the mean, which was the case for 35 of the 2560 recalls).

In addition to the correctness judgments, each recall video was also checked for possible delays in response times. During some trials, participants did not make a gesture immediately after releasing the spacebar but needed some more time to think. For these trials the correctness data was preserved, but the response time data was considered missing.

*Distraction task* - During the distraction task participants had to recall figure sequences of different lengths (see Appendix N). Of each recalled sequence the percentage of correct recall was calculated, as a measure of performance per participant. Only the sequences consisting of five or more figures were used to calculate the average percentage correct. If these average
scores were below twenty percent (more than three standard deviations from the mean), it was assumed that the participants were not focused enough on the distraction task to be distracted by the task. Therefore both recall and response time memorability scores (test 4) of these participants were excluded from the analysis. This was the case for six participants.

*Outliers* – One participant was detected as a potential outlier. This participant recalled on average only two gestures and recall decreased over time (which was opposite for other participants). The score for both average recall and improvement over trials was more than three standard deviations from the mean, and was deleted. The total of outliers and erroneous recordings, as described above, resulted in different number of participants between tests and dependent variables. The number of participant per test and measure is summarized by Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>106</td>
<td>85</td>
<td>78</td>
<td>69</td>
</tr>
<tr>
<td>Response time</td>
<td>74</td>
<td>66</td>
<td>61</td>
<td>55</td>
</tr>
</tbody>
</table>

*Analyses* – For both dependent variables two analyses were performed; one to test the effect on learnability (by using the average results of test 1, 2, and 3), and one to test the effect on memorability (by using the results of test 4, which is performed after the distraction task). The recall data were analyzed by means of logistic regression. The response time data were analyzed by means of a mixed model analysis. The analyses contained the same independent variables as used for the logistic regression models. The independent variables in the equations were teaching-method, intuitiveness, practice, gesture type, and test. In addition, it included the interaction effects of teaching-method and practice, teaching-method and gesture type, teaching-method and intuitiveness, and intuitiveness and gesture type. For the resulting general models for learnability and memorability see appendix Q. In addition, the effect of recall order, gender, the method-gender interaction, the method-test interaction, and the three-way method-intuitiveness-gesture type interaction was investigated, but no significant effects were found. These variables and interactions were not further considered during the analysis.
Despite careful selection of gestures, gesture had a significant effect on both response time (learnability: $F(8,195.86)=4.45$, $p<.01$, $r=.15$; memorability: $F(8,79.35)=3.95$, $p<.01$, $r=.22$), and recall (learnability: Wald $\chi^2(8)=29.27$, $p < .05$; on memorability: Wald $\chi^2(8)=12.29$, $p > .05$, see Figure 28 and Figure 29). These effects can be explained by the effects of gesture one and seven. These gestures could be interpreted in more than one way, as was described above, which could have attracted more attention, and therefore might improve recall for these gestures. Because the main effect of gesture type is of more interest to this experiment than the main effect of gesture, gesture was discarded from further analysis.

![Correct recall per gesture](chart1.png) ![Response time per gesture](chart2.png)

**Figure 28.** Percentage of correctly recalled gestures per gesture for all tests. Error bars indicate standard errors of the means.

**Figure 29.** Average response time per gesture, for all tests. Error bars indicate standard errors of the means.
3.2 Results

The description of the results is per effect divided in learnability results and memorability results. Learnability results are based on the combined results of test 1, 2, and 3, whereas memorability results are based on the results of test 4. Further explanation about this division can be found in the method section of experiment 2.

Main effects

Teaching method – On average, participants in the pantomime condition recalled fewer gestures (M = 61 percent, SE = 5 percent) during learnability tests than participants in the imitation (M = 69 percent, SE = 5 percent) or combination (M = 71 percent, SE = 4 percent) condition (see Figure 30). When taking into account the interaction effects, the main effect of teaching-method was significant (Wald χ²(2)=9.26, p < .05; imitation exp(b)=1.19, p > .05, pantomime exp(b) = 1.93, p < .01). In contrast to the effect on recall, a linear mixed model analysis revealed that there was no significant effect of teaching method on response time (F(2.56.58)=.46, p > .05) (see Figure 31).

The recall results for the different methods for the memorability test seem to disperse relative to the results of test 3. Recall in the pantomime condition decreases (to M = 77 percent, SE = 6 percent), while for the imitation and combination condition recall remains unchanged (imitation: M= 89 percent, SE = 4 percent, and combination: M = 92 percent, SE = 3 percent). The main effect of teaching method for memorability was significant (Wald χ²(2)= 8.93, p < .05). In contrast, the response time results after the distraction task did not show a significant effect of teaching-method (F(2.35.43)=0.34, p > .05).

Intuitiveness – As can be seen in Figure 32, recall was on average better for intuitive gestures compared to unintuitive gestures. For the learnability test (intuitive: M = 86 percent, SE = 2 percent, and unintuitive: M = 48 percent, SE = 3 percent) this effect was significant, and had a large effect size (Wald χ²(1)=59.33, p < .01, exp(b) = 4.973). In addition, during the learnability tests the average response time was smaller for the intuitive gestures (M = 2505 ms, SE = 1151 ms) compared to the unintuitive gestures (M = 4169 ms, SE = 1831 ms). A mixed model analysis
Figure 30. Percentage of correct recall per teaching-method and test. Error bars indicate standard errors of the means.

Figure 31. Response time in milliseconds per teaching-method and test. Error bars indicate standard errors of the means.

Figure 32. Percentage of correct recall per level of intuitiveness for all tests. Error bars indicate standard errors of the means.

Figure 33. Response time in milliseconds per level of intuitiveness for all tests. Error bars indicate standard errors of the means.

Figure 34. Average percentage of correct recall per practice condition and test. Error bars indicate standard errors of the means.

Figure 35. Average response time in milliseconds per practice condition and test. Error bars indicate standard errors of the means.
revealed that the effect on response time was significant and had a large effect size ($F(1,57.03)=27.00, p < .01, r = .57$).

When looking at the results measured after the distraction task (i.e. memorability), intuitive gestures were better recalled than unintuitive gestures (intuitive: $M = 98$ percent, SE = 0.7 percent, unintuitive: $M = 76$ percent, SE = 4 percent). The Wald statistic for this effect appeared insignificant ($\chi^2(1)=0.00, p > .05$). However, because the regression coefficient was large (19089) and the addition of intuitiveness to the regression model resulted in a significant improvement of the model ($\chi^2(1) = 17901, p < .01$) it can be assumed that the effect of intuitiveness on recall during the memorability test actually is significant. In addition, the effect on response time was significant (intuitive: $M = 1387$ ms, SE = 117 ms, unintuitive: $M = 2410$ ms, SE = 272 ms; $F(1,35.99)=13.82, p < .01, r =.53$) and had a large effect size.

**Practice** – Figure 34 shows the results on recall per practice condition. In contrast to what would be expected, recall was on average not better for participants in the practice condition (learnability tests: $M = 64$ percent, SE = 4 percent) compared to recall for participants in the no practice condition (learnability: $M = 69$ percent, SE = 4 percent). The effect of practice was not significant for both tests (learnability: Wald $\chi^2(1)=0.40, p > .05$, and memorability: Wald $\chi^2(1) = 0.46, p > .05$). Similar to recall, practice did not result in a shorter average response time (learnability: $M = 2679$ ms, SE = 252 ms) compared to the no practice condition (learnability: $M = 2862$ ms, SE = 283 ms, see Figure 35). The effect of practice was not significant for both tests (learnability: $F(1,56.93)=0.48, p > .05$, memorability: $F(1,35.54)=0.21, p > .05$).

**Gesture type** – To check whether gesture type influenced the results, both the main and the interaction effects of this control variable were analyzed. For both learnability and memorability recall for emblems was, on average, equal to recall for object-based gestures (emblems: $M = 72$ percent, SE = 3 percent, object-based gestures: $M = 67$ percent, SE = 3 percent; learnability: Wald $\chi^2(1) = 0.70, p > .05$, memorability: Wald $\chi^2(1) = 0.56, p > .05$). In contrast, the effect of gesture type on response time was significant for learnability although the effect size was small ($F(1,791.48)=6.95, p < .05, r=.09$). Object-based gestures resulted in a slightly shorter response times ($M = 2474$ ms, SE = 138 ms) compared to emblems ($M = 2597$ ms, SE = 200 ms). This effect was not found for the memorability test ($F(1,146.32)=0.00, p > .05$).
Interaction effects

Teaching-method*practice - Figure 38 shows the interaction between the teaching-methods and practice conditions. For learnability practice resulted on average in fewer correct recalls for the pantomime method (no practice: \(M = 62\) percent, \(SE = 8\) percent, and practice: \(M = 60\) percent, \(SE = 6\) percent), whereas it resulted on average in a similar percentage of correct recalls for the combination method (no practice: \(M = 73\) percent, \(SE = 4\) percent, and practice: \(M = 69\) percent, \(SE = 7\) percent). In addition, the overall interaction effect of teaching-method and practice was significant (Wald \(\chi^2(2) = 9.98\), \(p < .01\)). In addition, the interaction effect of teaching-method and practice on response time was significant (F(2,56.64) = 4.50, \(p < .05\), \(r = .27\); also see Figure 39). For the imitation and combination methods the results on response time were in line with what would be expected after seeing the recall results; practice resulted on average in similar response times compared to no practice (for the imitation method respectively \(M = 2695\) ms, \(SE = 281\) ms, and \(M = 2470\) ms, \(SE = 303\) ms; and for the combination method respectively \(M = 2925\) ms, \(SE = 499\) ms, and \(M = 2427\) ms, \(SE = 209\) ms). For the pantomime condition the results were not as what would be expected after seeing the effects on recall; practice resulted in smaller response times compared to no practice for the first test (respectively \(M = 2555\) ms, \(SE = 929\) ms, and \(M = 4578\) ms, \(SE = 833\) ms), but resulted in similar response times for test three (respectively \(M = 2682\) ms, \(SE = 931\) ms, and \(M = 2476\) ms, \(SE = 323\) ms).
In contrast to the results for the learnability tests, there was no significant interaction effect of teaching-method and practice on recall for the memorability test (Wald $\chi^2(2) = 0.25$, $p > .05$). In addition, there was no significant interaction effect on response time (F(2,35.29) = 0.81, $p > .05$).

**Intuitiveness, gesture type, & teaching-method** - Figure 40 and Figure 41 show the interaction effects between intuitiveness, gesture type, and teaching-method on correct recall and response time. The overall interactions of intuitiveness and gesture type on recall was not significant, but indicated a trend for learnability (learnability: Wald $\chi^2(1)$=3.72 , $p = .05$, exp(b) = 1.52; memorability: Wald $\chi^2(1)$=0.74 , $p > .05$). Emblems were on average recalled better for intuitive gestures (M = 90 percent, SE = 2 percent) compared to object-based gestures (M = 85 percent, SE = 3 percent). This difference is smaller for unintuitive gestures (respectively 54 percent, SE = 3 percent, and M = 49 percent, SE = 4 percent). In addition, there was no significant interaction effect on response time for the learnability tests (F(1,801.72)=.85, $p > .05$), but there was a significant effect for the memorability test (F(1,139.84)=5.07, $p < .05$, r = .19).

As can be seen in the figure, emblems were on average recalled equally well for all teaching-methods in the intuitive condition (approximately 90 percent), whereas the combination

\[43\]
condition resulted in more recall of emblems in the unintuitive conditions (combination method: \( M = 63 \) percent, \( SE = 5 \) percent, versus imitation method: \( M = 49 \) percent, \( SE = 6 \) percent, and pantomime method: \( M = 50 \) percent, \( SE = 6 \) percent). In the unintuitive condition object-based gestures were recalled less in the pantomime condition (\( M = 38 \) percent, \( SE = 6 \) percent) compared to the imitation and combination conditions (respectively \( M = 52 \) percent, \( SE = 7 \) percent, and \( M = 58 \) percent, \( SE = 6 \) percent). The difference between emblem and object-based gesture recall in the pantomime condition was largest for the unintuitive gestures. The interaction between teaching-method and gesture type was significant for the learnability test (Wald \( \chi^2(2)=10.91 \), \( p < .01 \)), but not for the memorability test (Wald \( \chi^2(2)=1.03 \), \( p > .05 \)). In addition, there were no significant effects on response time (learnability: \( F(2,840.56) = 1.12 \), \( p > .05 \); and memorability: \( F(2,142.53) = 0.71 \), \( p > .05 \)). The interaction between teaching-method and intuitiveness was significant for the learnability test (Wald \( \chi^2(2)=11.16 \), \( p < .01 \)), but not for the memorability test (Wald \( \chi^2(2)=0.17 \), \( p > .05 \)), and the response time tests (learnability: \( F(2,56.31) = 0.76 \), \( p > .05 \); memorability: \( F(2,35.20) = 0.26 \), \( p > .05 \)). No significant second-order interaction was found for method, intuitiveness and gesture type.
4. General discussion

During this graduation project the aim was to investigate two topics related to gesture-based human-technology interaction: intuitive gestures and gesture teaching-methods. This focus of both topics was relatively new in the field of gesture-based interfaces. The first experiment explored interaction gestures for controlling a television and focused on intuitive gestures that had a standalone meaning (so no direct manipulation was involved). The results showed many commonalities between improvised gestures of participants, especially for interaction tasks that were perceived as easy. These results are in line with the results of explorative studies focusing on intuitive touch-screen and object gestures (Wobbrock et al., 2009; Ruiz et al., 2011), suggesting that the type of gestures used in the current experiment is not intrinsically more difficult or more unintuitive. In addition, the association types were similar to what was found in the studies by Wobbrock et al. (2009) and Ruiz et al. (2011). Most gestures found in the current experiment originated from previous interactions with product interfaces and were pantomimes of these interactions, whereas the found touch-screen gestures were mainly based on the Windows desktop metaphor (Wobbrock et al., 2009) and the object gestures (for smartphone) were based on conventional phone-use. In our experiment pantomimes were mainly produced for easy tasks. The gestures that were produced for more difficult tasks originated more often from human interaction, and were mostly conventional gestures (emblems).

The results of this experiment suggest that for first time users the most intuitive gestures are probably either pantomime gestures or emblems. However, it should be noted that the use of intuitive gestures has advantages and disadvantages. The advantage might be that users learn these gestures quickly and understand the relationship between the gesture and the task. A disadvantage might be that these gestures are meaningful and therefore they might have large variation in form (Tessari et al., 2006). This might result in interpretation problems for gesture recognition systems. These problems could be solved by choosing meaningless gestures which are often produced with smaller variation. A further point of consideration, when selecting intuitive gestures is the association inherent to the gesture. Because intuitive gestures might function as a metaphor of a different product, the similarities between the interaction possibilities of the new product and the inferred interaction possibilities caused by the metaphor should be investigated. Otherwise this could lead to erroneous inferences and non-optimal interaction with
the product. This problem could especially occur for intuitive gestures that are based on the interaction with other products, but this might also occur for emblems.

This first experiment provided clear examples of intuitive gestures that could be used for the implementation in a gesture-based television interface. It should be stressed that the use of these gestures is almost certainly restricted to the context of the television system, because different types of devices might elicit different associations and therefore result in different gestures. In addition, it is recommended to redo the experiment at different points in time because the introduction of new technologies might result in different associations. A nice example of this is the swiping-movement that participants made for the ‘switching-channel’ tasks in this experiment. Participants who owned a smartphone or tablet PC reported to associate this movement with the interaction with a tablet or smart-phone, while participants who did not own such a device chose another gesture.

A selection of the results of the first experiment was used for the gesture-set used during the second experiment. In this experiment the effect of three gesture teaching-methods was tested on both learning speed and gesture memory. The effects were tested for varying conditions of gesture intuitiveness and practice. The results showed that, as expected, there was a large main effect of intuitiveness both on the percentage of correct recall and on response time. This finding stresses the importance of intuitive gestures for the learnability and memorability of a gesture-set.

The results showed that the effect of teaching-method was opposite to what was hypothesized; participants in the pantomime condition recalled, on average, fewer gestures than participants in the imitation and combination condition. The hypothesis was based on the view that observers are mainly interested in an actions goal (De Lange et al., 2008; Bekkering et al., 2000), and in the pantomime teaching-method the goal was explicitly visible. Participants in the combination condition recalled the largest percentage of gestures, which could suggest that the imitation method and the pantomime method complement each other. This might also imply that the combination condition resulted in more or stronger mental simulations, which would be in line with the view that the observation of direct manipulation results in a different type of simulation than the observation of pantomime or object movements individually (Cartmill et al., 2012). However, the results for the teaching-methods were highly influenced by the intuitiveness and the type of the gestures, and the practice condition participants were in. Within the intuitive condition, recall of emblems was for all methods approximately equal, whereas there were large
differences between teaching-methods in the unintuitive condition. The relatively high percentage of correct recalls in the intuitive condition might suggest a ceiling effect, but might also demonstrate certain properties of both teaching-method and gesture type. This latter view is further explained when focusing on the results other variables in the experiment. Recall of object-based gestures was lowest for the pantomime condition, especially for the unintuitive gestures. This is opposite to what was hypothesized. Although no convincing reason is available for this result at this point, it might be that the participants did not expect pantomimes, and this lack of expectation could have had different effects on the results per teaching-method. If this would be the case, a potential explanation for this might be that participants just imitate the hand movement in the imitation and combination condition without focusing on the gesture’s meaning, but experience more interpretation problems when they see only objects as movement-triggers. This interpretation problem might also explain the additional decrease in recall of unintuitive gestures for the pantomime method, compared to the other methods, because in this view the task association is more important to the pantomime method compared to the imitation and combination methods.

The results for ‘practice’ were also opposite to what was hypothesized. Practice did not improve recall as would be expected. In addition, practice resulted even in less recalls for the imitation and pantomime method. These findings are conflicting with the results of studies investigating the acquisition of dance movements (Cross et al., 2006; Thornton, & Knoblich, 2005), and are also not in line with the enactment paradigm which states that memory is better for tasks that are physically performed (Tessari et al., 2006). At this point no explanation is available for these results.

In total, three significant teaching-method interaction-effects on recall were found for the learnability tests. However, only the interaction with practice was found to have a significant effect on response time as well. In general, the response-time measure had a much lower power than the recall measure; only very strong effects (i.e. intuitiveness and practice-interaction) were reflected in the response time results. A lack of statistical power was also a problem for the memorability part of the experiment (which included the recall test after the distraction task). When participants changed their practicing behavior within the course of the experiment, their results of tests subsequent to the change were excluded, which resulted in a relatively small number of data points in the final recall test.
The overall results of this experiment shed light on some important issues when testing gesture teaching-methods. First, it shows that it is important to control for the intuitiveness of the gestures, because it has a large main effect on both recall and response time. Second, it shows that it is important to control for gesture type. Although gesture type did not have a significant main effect it did interact with the different teaching-methods, especially for unintuitive gestures. Third, it shows the importance of proper stimulus design for interpretation. Although the stimuli resulted from multiple iterations and were tested in preliminary tests, there were still participants who misinterpreted the meaning. And fourth, it is important to control the expectations of the participants related to the gestures, especially because the term ‘gesture’ can imply different types of movements. Because there was no control for this latter potentially confounding variable, the conclusions that can be drawn are limited.

Some important things related to the research questions of the second experiment can be concluded. First of all, all teaching methods can be used to teach gesture-sets. After three observations, participants correctly recalled on average more than eight out of ten gestures for all teaching-methods. This was the case for both participants who practiced and not practiced. Secondly, showing only a human arm or an arm and an object potentially results in the least amount of errors caused by interpretation and potentially by expectations. Thirdly, intuitive emblem gestures appear to be influenced less by the way they are taught, compared to object-based gestures. However, as mentioned before, this might be caused by the expectations that participants had about the types of gestures which were used in the experiment. Fourthly, for unintuitive gestures participants require many more stimulus presentations before remembering a complete gesture-set.

These first results suggest that in this situation both teaching-methods might complement each other. This would imply that the different teaching-methods affect different memory processes. According to Bartolo et al. (2003), pantomime gestures are formed within the working memory, and rely on stored information of perceptual object characteristics. For the pantomime teaching-method this would imply that the interface-object used for this teaching-method would be stored in memory. When object-based gestures are presented by means of the pantomime method, this would involve the same processes as just described; no additional processes are involved. However, when emblems are taught by the pantomime teaching-method, and the gesture is recognized, additional memory processes may be involved. This might explain the
differences in recall results found in the current experiment for emblems and object based gestures in the pantomime condition. During the imitation teaching-method participants observed hand movements. When participants in the imitation condition observed gestures three things could be possible. First, the hand movement could be recognized as a pantomime movement, and in this case (according to bartolo et al., 2003) the intended object would be stored in memory. Second, the gesture can be recognized and memorized as being an emblem, which would involve semantic processes that are also used for language processes (Willems, & Hagoort, 2007). Third, the gesture could also not be recognized, which makes it an unintuitive movement for which imitation involves processes as described by the dual-route model of imitation (Carmo, & Rumiati, 2009). To summarize, the results of the second experiment showed that recall was least affected for conditions in which both the means (the movement) and the goal of a gesture were communicated. This would suggest that although humans are assumed to be mainly interested in a movement’s goal during observational learning, the observed movements of gestures are processed and are assumed to significantly complement memory processes.

Although the second experiment showed some initially interesting results, no conclusive answer could be given to the first main question: Which method, the observation of a human movement or an object, supports learnability and memorability of interaction gestures most? However, all teaching-methods resulted in steep learning-curves, and therefore could be used to teach gesture-sets. In addition, it shows that the effect of practice is much smaller than was expected. To be able to answer the first main research question and to optimize the experiment, some suggestions to improve the current experiment will be given in the following paragraph.

There are potential improvements for different parts of the current experiment. The first part of the experiment that might be improved is the instruction. Participants should be told that the gestures they are going to see do not need to be related to the tasks and do not need to be conventional gestures. This adjustment should reduce potential problems caused by the expectations of participants, and would allow the experimenter to distinguish between recall errors caused by interpretation and memory problems. Also, the instructions could constrain the participants in the time they have to make the gesture. This is expected to increase the amount of useable response time data, because participants will probably hold the spacebar until they are more certain they know the answer. In addition, an intervention should be implemented in the instruction to make the practicing behavior of participants more consistent. This intervention
would make it possible to further investigate the counterintuitive effect of practice in the current experiment, but also makes it possible to investigate the effects of the teaching-methods on long-term recall. A second part of the experiment that might be improved is the set-up. Participants might be physically constrained to use their left hand, because several participants used both hands during the current experiment. The third part of the experiment that might be improved is the gesture-set. The video stimuli might be further improved to reduce the number of misinterpretations. In addition, the number of gestures in the gesture-set might be increased for the intuitive condition, to be able to measure a potential ceiling effect.

In addition to the current research questions, some suggestions for further research on gesture-based interaction are given next. First, it would be interesting to investigate the effect of stimulus variables such as gesture size and the level of detail involved, on the form and size of imitated gestures. In addition, it would be interesting to investigate if the form and size of these imitated interaction gestures changes over time by investigating longer-term use of gesture-based interfaces. Second, it would be interesting to look at the individual differences in gesture form. Definition of a user-specific ‘handwriting’ might aid gesture recognition in a similar way as it aids written-word recognition in the smartphone application described by Ouyang and Li (2012). In their study Ouyang and Li (2012) describe a word recognition system that feeds personal ‘handwriting’ characteristics back into the recognition system. In addition, when variations between participants are found to be large, the ‘handwriting’ information might even be used to identify users. A final suggestion for further research is related to the pantomime condition of the current research. In this condition participants observed animated moving interface objects that elicited gestures. It would be interesting to investigate if static icons could be designed that elicit similar movements as the animated interface objects. An advantage of the use of icons would be that they can be immediately interpreted by the users. This might aid recall, but might also have advantages for interface when these icons could be implemented as visual interaction cues for a gesture-based interface.

**Implications for design**

The current research project was performed in the context of a potential gesture-interface implementation in Philips televisions. The first experiment resulted in intuitive gestures for fourteen television control-tasks. When gestures need to be chosen for tasks that were not tested
in the current experiments, the method of the first experiment could be used to explore more intuitive gestures. Some general implications for the implementation of a gesture-based interface that resulted from this research project are described in the following paragraph.

To facilitate a steep learning-curve for the user, a gesture-set should be used that contains a maximum of eight gestures. Preferably, only ‘easy’ (as defined by the users) and common interaction-tasks should be performed by means of gestures. In addition, these gestures should be intuitive and should have a wide-spread task association. However, it should be noted that these intuitive gestures require a robust gesture-recognition system, because intuitive gestures have more variability in gesture form compared to unintuitive gestures that are not based on object-use or standard gestures. Intuitive gestures should be pantomimes of object-usage (of different products or interfaces) or be standardized gestures (emblems). These intuitive gestures might function as a metaphor. Because of this, the similarities between the interaction possibilities of the developed interface and the inferred interaction possibilities caused by the metaphor should be investigated to prevent erroneous expectations of the user.

All teaching-methods described in this report could be used to teach users a gesture-set containing intuitive gestures. For a gesture-set containing unintuitive gestures the combination method should be used. In addition, when direct and flawless interpretation is required, emblems must be used in stead of pantomime gestures. For all teaching-methods video-stimuli should be developed by means of an iterative user-centered design process, to ensure correct interpretation by the users. The teaching-methods might be implemented in a game or an other type of introduction task in which users are instructed to observe the video-stimuli and perform the communicated gestures. In addition, the users could be instructed to practice during the video observation.
5. References


Appendix A. Description scenarios

Scenario (use-case) descriptions

UC1. Sitting, easy interaction (‘watching a movie’)
After a day of work you planned to spend your evening watching some television shows. You are feeling tired and want a nice relaxing evening. You have settled yourself in a comfortable couch.

UC2. Standing, fast interaction (‘walking to the door’)
You have been watching a movie with a friend for a while, when suddenly the door-bell rings. You want to answer the door and you are walking to the hallway. Before you leave the room you quickly want to..

UC3. Sitting, working, remote (‘working’)
You are working at home and the television is on in the background. You are finishing the layout of a report that you want to email the next morning. In the meanwhile you pay some attention to what is on the television.

UC4. Standing, carrying, close (‘tidying the living room’)
You are tidying the living room because some friends will come and visit you in an hour. You already vacuumed the room, and now want to move some boxes to the room next door. While carrying a box you want to..
Appendix B. Questionnaire

Questionnaire

Please fill in the following questionnaire about your personal information, and your previous experiences with different types of interfaces.

Gender: □ Male   □ Female

Age: ________

Nationality: ______________________

Native language: ______________________

When you would compare yourself to others, what would you say about your experience with modern technology?
Extremely low □ □ □ □ □ extremely high

Do you use a smart-phone or a tablet with a touch-screen?
□ yes   □ no

When your answer to the previous question is yes, can you specify the type of device(s) you use and give an indication of how often you use it:

__________________________________________________________________________

__________________________________________________________________________

__________________________________________________________________________

Please turn this page for more questions..
Which statement about the Xbox Kinect, the Nintendo Wii, and the Playstation Move game consoles matches your experience best?

Xbox Kinect:
☐ I have never used it  ☐ I have used it before  ☐ I use it regularly  ☐ I use it often

Nintendo Wii:
☐ I have never used it  ☐ I have used it before  ☐ I use it regularly  ☐ I use it often

Playstation Move:
☐ I have never used it  ☐ I have used it before  ☐ I use it regularly  ☐ I use it often

Please write it down below if you have any other experience with devices that use body movements (gestures) as input:

____________________________________________________________________________________

____________________________________________________________________________________

Have you learned to communicate by using sign language (Dutch: gebarentaal)?
☐ yes  ☐ no

If you have any questions or remarks you can write it down or ask/tell the experimenter:

____________________________________________________________________________________

____________________________________________________________________________________

Thank you for your participation!
Appendix C. General instruction & Informed consent

Description of experiment

In the first part of this experiment you will be asked to think of and produce body movements that you think are suitable for the specific tasks and contexts. All tasks are about changing television settings.

The second part of the experiment will consist of an interview and a questionnaire. You are being filmed during both parts of the experiment.

The whole experiment will last approximately 1 hour. Participation in this experiment is strictly voluntary. You may choose to stop participation and withdraw from this study at any time, without providing reason. Please indicate the researcher if you wish to stop.

Informed consent

Title of experiment: TV taak

Responsible researcher: Eline Jansen

To be filled in by the participant:

I acknowledge that I have read this consent form and I understand its content. I understand that all the data resulting from this experiment will be handled confidentially, and my name will not be included, or in any other way associated, with the data collected in this study. All (video) data is used only for analysis and scientific purposes related to the current study.

I voluntarily participate in this research study. I acknowledge that I may stop participation at any time, without providing any reason.

Name
participant: …………………………………………………………………………………………

Date: ……………………….. Signature participant: ………………………………
Appendix D. Set-up experiment

Figure 1 shows a schematic representation of the set-up. The characters refer to:

A: Position camera 1
B: Position camera 2
1: Location participant for use-case 'tidying the living room'
2: Location participant for use-case 'walking to the door'
3: Location participant for use-case 'watching a movie'
4: Location participant for use-case 'working'

Figure 42. Schematic overview set-up. Numbers indicate participants position during scenarios, A and B indicate camera positions.
## Appendix E. Description variables data-file

Table 4. Elaborate description of the variables in the dataset of experiment 1

<table>
<thead>
<tr>
<th>Column</th>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Use-case</td>
<td>Scenario condition: Sit = UC1, Door = UC2, Work = UC3, Box = UC4</td>
</tr>
<tr>
<td>B</td>
<td>Task</td>
<td>Task condition: name of task</td>
</tr>
<tr>
<td>C</td>
<td>Pp no.</td>
<td>Participant number</td>
</tr>
<tr>
<td>D</td>
<td>Video file name</td>
<td>Video file name</td>
</tr>
<tr>
<td>E</td>
<td>Link</td>
<td>Link to video file</td>
</tr>
<tr>
<td>F</td>
<td>Association</td>
<td>Literal association of participant</td>
</tr>
<tr>
<td>I</td>
<td>Relative to other gestures/tasks (not related/else)</td>
<td>How current task is related to related task (e.g. switch on and off)</td>
</tr>
<tr>
<td>J</td>
<td>additional info</td>
<td>Additional comments of participants that are related to task</td>
</tr>
<tr>
<td>K</td>
<td>most suitable</td>
<td>If the gesture was mentioned to be the most suitable gesture, then ‘x’</td>
</tr>
<tr>
<td>L</td>
<td>interpretation</td>
<td>Categorization based on both association and form (1 observer)</td>
</tr>
<tr>
<td>M</td>
<td>Arm movement</td>
<td>Yes if arm is moved, no if not</td>
</tr>
<tr>
<td>N</td>
<td>single/both hands</td>
<td>Number of arms used to gesture</td>
</tr>
<tr>
<td>O</td>
<td>contacting objects/body parts</td>
<td>If gesture involved contacting objects or body parts, then ‘y’</td>
</tr>
<tr>
<td>P</td>
<td>general direction (horizontal/vertical/diagonal/circular/static/z/other)</td>
<td>Orientation of overall gesture (no attention for details)</td>
</tr>
<tr>
<td>Q</td>
<td>specific direction (left, right, up, down, to the front, back)</td>
<td>Direction of gesture (if one direction could be specified)</td>
</tr>
<tr>
<td>R</td>
<td>stepwise (only video-task) (if stepwise describe per step) (1, 2, multiple steps)</td>
<td>If no fluent movement was made but a stepwise gesture instead (only for video task)</td>
</tr>
<tr>
<td>Column</td>
<td>Title</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>S</td>
<td>z-axis</td>
<td>Did the gesture include a movement in the z-direction?</td>
</tr>
<tr>
<td>T</td>
<td>repetitive</td>
<td>Was a single movement repeated?</td>
</tr>
<tr>
<td>U</td>
<td>speed (constant/varying)</td>
<td>Was the gesture performed in a constant or fluctuating speed?</td>
</tr>
<tr>
<td>V</td>
<td>Upper arm movement</td>
<td>Did the gesture contain a movement of the upper arm?</td>
</tr>
<tr>
<td>W</td>
<td>direction (yaw/roll/pitch)</td>
<td>Orientation of movement.</td>
</tr>
<tr>
<td>X</td>
<td>trajectory (linear/curved linear/circular)</td>
<td>Trajectory of movement</td>
</tr>
<tr>
<td>Y</td>
<td>elbow (bent/straight)</td>
<td>Elbow position</td>
</tr>
<tr>
<td>Z</td>
<td>Lower arm movement</td>
<td>Did the gesture contain a movement of the lower arm?</td>
</tr>
<tr>
<td>AA</td>
<td>direction (horizontal/vertical/diagonal/circular/z)</td>
<td>Orientation of movement</td>
</tr>
<tr>
<td>AB</td>
<td>trajectory linear/curved linear/circular</td>
<td>Trajectory of movement</td>
</tr>
<tr>
<td>AC</td>
<td>Wrist movement</td>
<td>Did the gesture contain a movement of the wrist?</td>
</tr>
<tr>
<td>AD</td>
<td>movement direction (horizontal/vertical/diagonal/circular/z)</td>
<td>Orientation of movement</td>
</tr>
<tr>
<td>AE</td>
<td>position (stretched/bent)</td>
<td>Position of wrist</td>
</tr>
<tr>
<td>AF</td>
<td>Hand movement</td>
<td>Did the gesture contain a hand movement</td>
</tr>
<tr>
<td>AG</td>
<td>description (opening/closing/snapping fingers/else)</td>
<td>General description of hand movement</td>
</tr>
<tr>
<td>AH</td>
<td>palm (l) facing (0-360 tv-face)</td>
<td>Orientation of left hand</td>
</tr>
<tr>
<td>AI</td>
<td>palm (r) facing (0-360 tv-face)</td>
<td>Orientation of right hand</td>
</tr>
<tr>
<td>AJ</td>
<td>form (starting) (open/spread/fist/rknob full grip/rknop prec grip/else)</td>
<td>Position of hand at the beginning of a movement</td>
</tr>
<tr>
<td>AK</td>
<td>thumb (straight/bent)</td>
<td>Position of thumb</td>
</tr>
<tr>
<td>AL</td>
<td>index (straight/bent)</td>
<td>Position of index finger</td>
</tr>
<tr>
<td>AM</td>
<td>middle (straight/bent)</td>
<td>Position of middle finger</td>
</tr>
<tr>
<td>AN</td>
<td>ring (straight/bent)</td>
<td>Position of ring finger</td>
</tr>
<tr>
<td>AO</td>
<td>little (straight/bent)</td>
<td>Position of little finger</td>
</tr>
<tr>
<td>Column</td>
<td>Title</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>AP</td>
<td>finger movement</td>
<td>Does the gesture include a movement of a finger?</td>
</tr>
<tr>
<td>AQ</td>
<td>thumb movement</td>
<td>Does the gesture include a movement of the thumb?</td>
</tr>
<tr>
<td>AR</td>
<td>index finger movement</td>
<td>Does the gesture include a movement of the index finger?</td>
</tr>
<tr>
<td>AS</td>
<td>middle finger movement</td>
<td>Does the gesture include a movement of the middle finger?</td>
</tr>
<tr>
<td>AT</td>
<td>ring finger movement</td>
<td>Does the gesture include a movement of the ring finger?</td>
</tr>
<tr>
<td>AU</td>
<td>little finger movement</td>
<td>Does the gesture include a movement of the little finger?</td>
</tr>
<tr>
<td>AV</td>
<td>Leg movement</td>
<td>Does the gesture include a movement of the leg?</td>
</tr>
<tr>
<td>AW</td>
<td>Arm movement</td>
<td>Does the gesture include a movement of the head?</td>
</tr>
</tbody>
</table>
### Appendix F. Gesture commonalities

Table 5. Summary of most used gesture categories per task.

<table>
<thead>
<tr>
<th>Task</th>
<th>First pattern</th>
<th>Second pattern</th>
<th>Third pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Switch on</strong></td>
<td>Snapping fingers</td>
<td>Opening a hand</td>
<td></td>
</tr>
<tr>
<td><strong>Switch off</strong></td>
<td>Horizontal arm movement</td>
<td>Show an open hand</td>
<td>Snapping fingers</td>
</tr>
<tr>
<td><strong>Increase the volume a little</strong></td>
<td>Arm movement upwards</td>
<td>Rotary knob</td>
<td>Show open hand</td>
</tr>
<tr>
<td><strong>Increase the volume much</strong></td>
<td>Arm movement upwards</td>
<td>Rotary knob</td>
<td>Pointing up</td>
</tr>
<tr>
<td><strong>Decrease the volume a little</strong></td>
<td>Arm movement downwards</td>
<td>Rotary knob</td>
<td>Show open hand</td>
</tr>
<tr>
<td><strong>Decrease the volume much</strong></td>
<td>Arm movement downwards</td>
<td>Rotary knob</td>
<td></td>
</tr>
<tr>
<td><strong>Mute</strong></td>
<td>Closing a hand</td>
<td>Blocking ears</td>
<td></td>
</tr>
<tr>
<td><strong>Un-mute</strong></td>
<td>Opening hand</td>
<td>Un-blocking ears</td>
<td></td>
</tr>
<tr>
<td><strong>Switch to the next channel</strong></td>
<td>Swiping</td>
<td>Pointing</td>
<td></td>
</tr>
<tr>
<td><strong>Switch to the previous channel</strong></td>
<td>Swiping</td>
<td>Pointing</td>
<td></td>
</tr>
<tr>
<td><strong>Activate the gesture interface</strong></td>
<td>Waving hands</td>
<td>Snapping fingers</td>
<td>Clapping hands</td>
</tr>
<tr>
<td><strong>De-activate the gesture interface</strong></td>
<td>Form a cross</td>
<td>Clapping hands</td>
<td></td>
</tr>
<tr>
<td><strong>Select the video file</strong></td>
<td>Pointing</td>
<td>Pantomime of steps</td>
<td></td>
</tr>
<tr>
<td><strong>Open the video file</strong></td>
<td>Pressing</td>
<td>Remaining position</td>
<td></td>
</tr>
</tbody>
</table>
Appendix G. Preliminary tests

Table of contents

Introduction........................................................................................................................................ 66
Test 1: stimuli categorization ........................................................................................................... 66
  Method........................................................................................................................................... 69
  Results........................................................................................................................................... 71
Test 2: stimuli categorization: iteration 1 .......................................................................................... 75
  Method........................................................................................................................................... 75
  Results........................................................................................................................................... 79
Discussion & choice for stimuli......................................................................................................... 82

Introduction

In this report experiments are described in which potential stimuli, to be used for the second experiment, have been tested and adjusted. The goal of this experiment was fourfold. First, the experiment should test the clarity of the individual stimuli. Second, the experiment should test the relative differences in recall difficulty between stimuli. These first two points where intended to aid the choice for stimuli. Third, the experiment should result in an estimation of the number of gestures participants could remember after one, two, or three exposures, and if there were preliminary differences between the methods of presentation: observing an object versus a human arm. Fourth, the experiment should result in an estimation of the duration of the second experiment. The complete report contains two experiments which will be described in chronological order in the following paragraphs.

Test 1. Stimuli categorization

In test 1 the clarity and memorability of fourteen pantomime stimuli and thirteen imitation stimuli were tested (for an overview see Figure 43). The stimuli’s variability on gesture origin and form was large. Intuitive gestures were chosen from the results of experiment one, and stimuli were designed to elicit the movements inherent to these gestures. Both object-based gestures and emblems (or quotable gestures) were selected (see caption Figure 43), and for each selected gesture at least one pantomime stimulus, showing an object, and one imitation stimulus,
showing a human arm, was designed. Four intuitive gestures (gesture B, H, J, & L) were intended for the unintuitive group by potentially pairing them with a non intuitive task. Two gestures were invented to make this set complete (gesture A & I).

During the experiment participants were exposed to either pantomime stimuli or imitation stimuli. The experiment was divided in three parts. In the first part participants were exposed to all stimuli sequentially, and were asked to perform a matching gesture after each stimulus. In the second part participants were asked to remember sequences of video stimuli in which each stimulus was paired with a number. And in a subsequent recall task participants had to recall the gesture that belonged to each number in random order. In the third part participants reviewed each stimulus and were asked if they had any associations with the stimuli.
Figure 43. Stimuli overview preliminary test 1; stimuli for the pantomime condition (upper half) and stimuli for the imitation condition (bottom half). Stimuli C, E, F, J, K, and L communicate object-based gestures and stimuli A, B, D, G, H, J, M, and N not object-based gestures.
Method

Design
Stimulus type (pantomime or imitation stimulus) was manipulated between participants, and gesture type (intuitive/not intuitive and object-based/not object-based) was varied within participants. During the first and third part of the experiment participants observed the stimuli of all gestures. During the second part of the experiment a random selection was made of the stimuli, to match the number of gestures participants had to recall.

The dependent variable of the first and second part of the experiment was correct gesture form which could be judged as correct or incorrect. For the criteria that were used to make this judgment see Table 6. The other dependent variable in the second part of the experiment was the number of stimuli that participants correctly recalled after one, two, three, and four times of stimuli observation. In addition the duration of this procedure was recorded for each participant. The dependent variable of the third part of the experiment was the reported association.

Participants
Ten participants participated in this experiment; three females (average age 27, ranging from 24 to 31 years old) and seven males (average age 25, ranging from 23 to 30 years old). All participants had normal or corrected to normal vision and were all students or employees working at Philips BG TV Innovation Site Eindhoven. The participation was completely voluntary, and the participants did not receive an incentive for their participation.

Table 6. Criteria for the ‘correct gesture form’ of test 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Judgment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Rotating hand</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>B</td>
<td>Semi circle</td>
<td>Upwards curved vertical movement</td>
</tr>
<tr>
<td>C</td>
<td>Push button</td>
<td>Movement hand towards the screen</td>
</tr>
<tr>
<td>D</td>
<td>Closing hand</td>
<td>Hand first opened, then closed</td>
</tr>
<tr>
<td>E</td>
<td>Remote 1</td>
<td>Palm directed to left or right, combined with a vertical thumb movement</td>
</tr>
<tr>
<td>F</td>
<td>Remote 2</td>
<td>Palm directed to left or right, combined with a vertical thumb movement</td>
</tr>
<tr>
<td>G</td>
<td>Showing hand</td>
<td>Opened had with palms facing screen</td>
</tr>
<tr>
<td>H</td>
<td>Diagonal slider</td>
<td>Upwards directed wrist or arm movement moving diagonal left or right</td>
</tr>
<tr>
<td>I</td>
<td>Draw square</td>
<td>Vertical movement, changing direction by making two abrupt curved movements</td>
</tr>
<tr>
<td>J</td>
<td>Spread fingers</td>
<td>Thumb and one other finger moving away from each other in a vertical manner</td>
</tr>
<tr>
<td>K</td>
<td>Rotary knob</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>L</td>
<td>Vertical slider</td>
<td>Vertical movement of arm or wrist</td>
</tr>
<tr>
<td>M</td>
<td>Wave one time</td>
<td>Horizontal movement changing direction once</td>
</tr>
<tr>
<td>N</td>
<td>Wave two times</td>
<td>Horizontal movement changing direction twice</td>
</tr>
</tbody>
</table>
Stimuli, apparatus, & Set-up

The stimuli were either video recordings of a hand making a gesture (in the imitation condition), or were animations of interface objects that were designed to elicit the gesture’s movement (in the pantomime condition). The stimuli were made in the same way and had similar characteristics as the stimuli of the second experiment. The stimuli were presented full screen on a 15 inch screen (resolution 1280x1024) that was positioned at 2.5 meters from the participant at a height of 1.5 meters.

For the first and third part of the experiment a Macromedia Authorware software program was made, that was able to initiate a stimulus presentation from a remote location by the experimenter. During the second part of the experiment pre-edited stimulus sequences were played by using VLC media player software. Each stimulus in the sequence was paired with a task number and between each video a blank screen was shown for one second.

The arm movements of the participants were recorded during part one and two of the experiment by means of a high definition camera that was positioned at the right side of the screen. The experimenter was in the room to test the recall and to judge in real-time if the recalled gestures were correct.

Procedure

After arrival participants were informed about the general procedure of the experiment and were asked for their informed consent (see appendix H). After this, the procedure of the first part of the test was explained in more detail by a text on the screen (see appendix I for the instruction texts used). The participant were informed that they were allowed to ask questions when something was not clear and were instructed to stand at a predefined location in front of the television. When the participants said they understood the instructions, and were ready, the first part of the experiment was started.

In the first part of the experiment the participants were presented with movie clips showing fourteen stimuli in the pantomime condition or thirteen stimuli in the imitation condition. After each movie clip (showing a single stimulus) the participants were asked to make the corresponding gesture, and were asked to judge the difficulty of the recall tasks on a five-point scale ranging from ‘very easy’ to ‘very difficult’. The participants did not receive feedback about the correctness of the gestures. This procedure was repeated for all stimuli.
After finishing the first part of the experiment, the participants were presented with the instruction of the second part of the experiment, in which they were instructed to remember which gesture was matched with which task number. When the participants confirmed having understood the task they were presented with a sequence of eight video-stimuli. Subsequently, the experimenter asked the participants to recall the gestures for each task in random order. The participants viewed and recalled the stimuli until they were able to recall all gestures.

In the third part of the experiment participants were exposed to the same stimuli as they had seen in the first part of the experiment. Instead of having to produce the matching gestures they were asked to report associations they had with the inherent gesture. After having finished this final part of the experiment they were thanked for their participation.

Results

1. Clarity of stimuli

As can be seen in Figure 44, almost all (12 out of 14) gestures were performed correctly by all participants in the imitation condition (n=5). The gesture stimuli that did not result in a correct answer for all participants were only misinterpreted by one out of five participants. Gesture stimuli in the object group were much more vulnerable to misinterpretation. Half of the stimuli were misinterpreted by at least one participant.

![Total correct gesture performance per task](image)

Figure 44. Total correct gesture performance per task. Each group contained five participants. For an overview of the stimuli see Figure 43.
Figure 45 shows a second measure of the clarity of the stimuli. Participants were asked to report on a five-point Likert scale ranging from ‘very easy’ to ‘very difficult’ to indicate how difficult they thought the task was. For the imitation group this implied the difficulty to imitate the gesture they observed when viewing the gesture-stimuli; for the object group it implied the ability to think of a matching gesture. In both the object and imitation group it was not controlled for the difficulty of the arm movement.

When comparing the results in Figure 44 and Figure 45 it appears that for approximately half of the gestures the perceived difficult can predict the performance. This lack of consistency might be due to the inherent perception of movement complexity and the perceived transitivity of the movement.

![Average perceived difficulty per task (n=10)](image)

Figure 45. Perceived difficulty per task for object and imitation group. Answers were given on a five-point Likert scale ranging from ‘very easy’ to ‘very difficult’. Error bars indicate standard errors of the mean. Each group contained 5 participants. For an overview of the stimuli see Figure 43.

2. Recall difficulty per stimulus

Figure 46 showed the average required viewings before correct recall was achieved per gesture. The variation is relatively low. This might be due to the number of stimuli in the memory task (eight) which made the task relatively simple. However, some gestures appear more difficult to recall for one group than the other (i.e. gesture D and B). These differences should be reduced to be able to measure an effect of teaching-method in the main experiment.
Figure 46. Average required viewings before correct recall. Gestures are divided into four groups based on gesture type; I&OB: Intuitive and Object-Based, I&NOB: Intuitive and Not Object-Based, NI&OB: Not Intuitive and Object-Based, and NI&NOB: Not Intuitive and Not Object-Based.

3. Memory span for gestures

In the current experiment sequences of eight gesture stimuli were used during the recall task. As a result, participants required relatively few observations before they were able to recall all gestures. As can be seen in Figure 47, the largest group (4 out of 10) recalled all gestures after viewing the sequence two times. In addition, the learning curves of the participants were relatively steep, as is shown by Figure 48.

Figure 47. Required viewings of gesture stimuli before complete recall. Maximal number of viewings was four. For the data in the ‘more’ column it is assumed that participants would recall all gestures after subsequent viewings.
Figure 48. Gesture learning-curve per participants. The total number of gestures to be recalled was eight.
Test 2. Stimuli categorization: iteration 1

In test 2 the clarity and memorability of thirteen new or revised pantomime stimuli and eleven imitation stimuli was tested. The stimuli in this test were all potential intuitive gestures, and were based on the results from pretest 1 and on standard emblem gestures (for an overview see Figure 49). The procedure was exactly the same as in test 1 except for part two in which more stimuli were added to each sequence.

Method

Design

The design of the experiment was similar to that of test 1 except from the fact that only potential intuitive gesture stimuli were tested. For the set of gestures that were tested and their characteristics see Figure 49.

Participants

Eight participants participated in this second experiment; two females (22 and 26 years old) and six males (average age 26, ranging from 23 to 34 years old). All participants had normal or corrected to normal vision and were all students or employees working at Philips BG TV Innovation Site Eindhoven. None of the participants had participated in the first preliminary test. Participation was completely voluntary, and the participants did not receive an incentive for their participation.

Table 7. Criteria for correct gesture form of test 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Judgment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Rotary knob 1</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>B</td>
<td>Rotary knob 2</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>C</td>
<td>Remote 1</td>
<td>Palm directed to left or right, combined with a vertical thumb movement</td>
</tr>
<tr>
<td>D</td>
<td>Remote 2</td>
<td>Palm directed to left or right, combined with a vertical thumb movement</td>
</tr>
<tr>
<td>E</td>
<td>Push button</td>
<td>Movement hand towards the screen</td>
</tr>
<tr>
<td>F</td>
<td>Swipe</td>
<td>Moving hand, wrist, or arm in vertical position</td>
</tr>
<tr>
<td>G</td>
<td>Zoom in</td>
<td>Spreading thumb and one other finger in a diagonal orientation</td>
</tr>
<tr>
<td>H</td>
<td>Wave</td>
<td>Horizontal movement changing direction once</td>
</tr>
<tr>
<td>I</td>
<td>Show hand</td>
<td>Opened had with palms facing screen</td>
</tr>
<tr>
<td>J</td>
<td>Close hand</td>
<td>Hand first opened, then closed</td>
</tr>
<tr>
<td>K</td>
<td>Close fingers</td>
<td>Bringing thumb and one other finger closer together in a vertical orientation</td>
</tr>
<tr>
<td>L</td>
<td>Previous</td>
<td>Upwards-curved vertical movement</td>
</tr>
<tr>
<td>M</td>
<td>Number two</td>
<td>Stretching out index finger and middle finger, while bending ring and little finger</td>
</tr>
</tbody>
</table>
Stimuli, apparatus & set-up
For an overview of the characteristics of the stimuli that were tested in this test see Figure 49. All other aspects related to the stimuli, apparatus and set-up are similar to test 1.

Procedure
The procedure was similar as in test 1, except from the instruction of part one and two, in which participants were instructed only to use one hand to make gestures (see appendix I). This was done because some variation occurred in test 1 because some participants tend to use both hands when possible.
Figure 49. Stimuli overview preliminary test 2; stimuli for the pantomime condition (upper half) and stimuli for the imitation condition (bottom half). Stimuli for object-based gestures are shown at the left and stimuli for not object-based gestures at the right.
Results

1. Clarity of stimuli

As can be seen in Figure 50 most stimuli were being understood properly. Only in five cases an incorrect gesture was made: three times in the object group and two times in the imitation group.

![Total correct gesture performance per task (n=8)](image)

Figure 50. Total correct gesture performance per task. Each group contained 4 participants. For an overview of the stimuli see Figure 49.

Figure 51 shows the perceived difficulty of the tasks that participants indicated on a five-point Likert scale ranging from ‘very easy to very difficult’. Two main differences between the object and imitation group were visible. The average perceived difficulty of task I in the object group was two points higher than in the imitation group. However, this is not reflected in the gesture performance (see Figure 50). Also, task M is judged as significantly more difficult in the object group (by more than two points). This might be reflected by misinterpretation of one participant as is visible in Figure 50.

When comparing the perceived difficulty of task I and M to Figure 52, which indicates the difficulty to recall the gestures, no effect on recall difficulty was found.
Perceived difficulty per task (n =8)

Figure 51. Perceived difficulty per task for object and imitation group. Answers were given on a five-point Likert scale ranging from ‘very easy’ to ‘very difficult’. Error bars indicate standard errors of the mean. Each group contained 4 participants. For an overview of the tasks see Figure 49.

2. Recall difficulty per stimulus

Figure 52 shows the average required stimulus viewing-times before participants recalled the gesture correctly. As can be seen in the figure the variability is relatively low; only two of the eleven gestures fall require on average more than two viewings before being recalled (C and H).

Average required viewings before correct recall per gesture (n=8)

Figure 52. Average required viewings before correct recall for object-based and not object-based gestures.
3. Memory span for gestures

Figure 53 shows the average required viewing times before correct recall. It appears as if there is a trend that participants in the object group are faster in recalling all gestures compared to the imitation group. Similar to the results of test 1, the minimal amount of viewing times to recall all gestures is two.

![Viewing times required for complete recall (n=8)](image)

Figure 53. Required viewings of gesture stimuli before complete recall. Maximal number of viewings was four.

Figure 54 shows the gesture learning curve per participant. The curves of the participants in the object group (right part of graph) appear to be slightly steeper, which would indicate a better memory.
Figure 54. Gesture learning curve per participants. Participants 1 and 2 were required to recall 9 gestures; all other participants had to recall 10 gestures.


Discussion

In the first preliminary test fourteen ‘pantomime’ and thirteen ‘imitation’ stimuli were tested. These stimuli were first version custom-made stimuli that were not based on theories or guidelines. The first results showed that properly designed ‘pantomime’ and ‘imitation’ stimuli can elicit predefined gestures. In addition, the results showed that participants’ perceived difficulty of a task could vary much while the required viewing time before correct recall did not vary. Therefore most weight was given to the ‘required viewing time’ variable during the choice for gestures. The results indicated also a number of problems, especially with some of the ‘pantomime’ stimuli. The problematic stimuli were replaced or adjusted which led to the new set of gestures that was tested in preliminary test 2 (see Figure 49 for an overview).

In the second preliminary test twelve ‘pantomime’ and eleven ‘imitation’ stimuli were tested. Results showed improved results on all measurements compared to the results of the first test. In addition, it showed that the memory span for gestures was relatively large. After three observations participants could recall ten gestures that were paired with an unintuitive name. The recall results suggested that ‘pantomime’ gestures might be recalled faster than ‘imitation’ gestures. The results for the individual stimuli showed that in theory all gestures of test 2 might be suitable to use in the main experiment, because the ‘pantomime’ and ‘imitation’ gestures require a similar amount of viewing times before correct recall, and are being interpreted as the same movement. However, gesture G was identified as a potential outlier because there was a relatively large difference in required viewing times between the different stimuli. In addition, gestures A, B, K, and M were misinterpreted by at least one participant. Gestures B, D, E, F, G, H, I, J, L, and M were selected for the main experiment. Some small adjustments were made to the stimuli to further improve interpretation before the gestures were used. Finally, the preliminary tests resulted in an duration estimation of thirty minutes for the main experiment.
Appendix H. Informed consent preliminary tests

Description of experiment

In the first part of the experiment you will be asked to make hand and/or arm movements after viewing short movie clips.

The second part of the experiment includes a number of memory tasks in which you will be asked to remember a sequence of gestures. You are being filmed during both parts of the experiment.

The whole experiment will last approximately 30 minutes. Participation in this experiment is strictly voluntary. You may choose to stop participation and withdraw from this study at any time, without providing a reason. Please indicate the researcher if you wish to stop.

Informed consent

Title of experiment: Pretest gesture tasks

Responsible researcher: Eline Jansen

To be filled in by the participant:

I acknowledge that I have read this consent form and I understand its content. I understand that all the data resulting from this experiment will be handled confidentially, and my name will not be included, or in any other way associated, with the data collected in this study. All (video) data is used only for analysis and scientific purposes related to the current study.

I voluntarily participate in this research study. I acknowledge that I may stop participation at any time, without providing any reason.

Name participant: ..............................................................................................................................................

Date: ........................................ Signature participant: .................................................................
Appendix I. Instructions preliminary tests

Table 8 shows the instructions per part of the experiment and per condition. The underlined instructions differed between tests. Only in the second test participants were instructed that all gestures only required one hand.

Table 8. Instructions per part of the experiment per experimental condition.

<table>
<thead>
<tr>
<th>Part of experiment</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part 1</strong>&lt;br&gt;Pantomime group</td>
<td>In the first part of this experiment you will be presented with 14 movie clips. Each movie clip shows an interface object which could be associated with a gesture (hand/arm movement). When the experiment starts you will be presented with one interface object at a time. When you know a gesture that you think suits the presented interface object, please make this gesture, directed towards the monitor. <strong>All gestures require only one hand.</strong> Please tell the experimenter when this instruction is not clear to you. When everything is clear the first part of the experiment will start.</td>
</tr>
<tr>
<td><strong>Part 1</strong>&lt;br&gt;Imitation group</td>
<td>In the first part of this experiment you will be presented with 13 movie clips. Each movie clip shows a human arm making a gesture (hand / arm movement). When the experiment starts you will be presented with one arm movement at a time. Your task is to imitate each movement. When you know which movement you have to make, please make this gesture, directed towards the monitor. <strong>All gestures require only one hand.</strong> Please tell the experimenter when this instruction is not clear to you. When everything is clear the first part of the experiment will start.</td>
</tr>
<tr>
<td><strong>Part 2</strong>&lt;br&gt;Both groups</td>
<td>In the second part of this experiment you will repeatedly view a sequence of movie clips. The movie clips are a selection of the movies you’ve watched in the first part of the experiment. In this part of the experiment the movie clips are paired with a number. When viewing the sequence, your task is to remember the gesture that is paired with each number. After viewing the sequence you will be asked to recall the gestures that you’ve seen in the movie clips. During the recall task you will be given only the number, and you have to recall the correct gesture. The recall task will be in random order. When everything is clear the second part of the experiment will start in a minute.</td>
</tr>
<tr>
<td><strong>Part 3</strong>&lt;br&gt;Both groups</td>
<td>In the last part of the experiment you will view all movie clips again. Please report any associations you have with the object movements.</td>
</tr>
<tr>
<td><strong>End</strong>&lt;br&gt;Both groups</td>
<td>This was the last task. Thank you for your participation!</td>
</tr>
</tbody>
</table>
## Appendix J. Overview stimuli

<table>
<thead>
<tr>
<th>Gesture number</th>
<th>Imitation</th>
<th>Pantomime</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Intuitive name: Decrease volume</td>
<td>Description: rotate rotary knob</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Select</td>
<td>Duration: 2 seconds</td>
</tr>
<tr>
<td></td>
<td>Gesture type: pantomime</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Intuitive name: Turn off</td>
<td>Description: press button remote-control</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Zoom in</td>
<td>Duration: 1.5 seconds</td>
</tr>
<tr>
<td>2</td>
<td>Intuitive name: Select</td>
<td>Description: press button</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Decrease volume</td>
<td>Duration: 1 second</td>
</tr>
<tr>
<td></td>
<td>Intuitive name</td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>3</td>
<td>Next</td>
<td>swipe</td>
</tr>
<tr>
<td>4</td>
<td>Zoom in</td>
<td>Touch-screen zoom</td>
</tr>
<tr>
<td>5</td>
<td>Pay attention</td>
<td>wave</td>
</tr>
<tr>
<td></td>
<td>Intuitive name: Stop</td>
<td>Description: show hand</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Previous</td>
<td>Duration: 2 seconds</td>
</tr>
<tr>
<td>6</td>
<td>Intuitive name: Mute</td>
<td>Description: close hand</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Number two</td>
<td>Duration: 1.5 seconds</td>
</tr>
<tr>
<td>7</td>
<td>Intuitive name: Previous</td>
<td>Description: draw semi-circle</td>
</tr>
<tr>
<td></td>
<td>Unintuitive name: Pay attention</td>
<td>Duration: 1 second</td>
</tr>
<tr>
<td>Intuitive name: Number two</td>
<td>Description: Show two fingers</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td>Unintuitive name: Next</td>
<td>Duration: 2 seconds</td>
<td></td>
</tr>
</tbody>
</table>

Figure 55. Overview of imitation and pantomime stimuli with inherent task-associations (intuitive & unintuitive name), and stimulus characteristics.
Appendix K. Criteria for recall judgments

Criteria list

Table 9. Original criteria list experiment 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Judgment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Rotary knob</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>1</td>
<td>Remote</td>
<td>Palm directed to left or right, combined with a vertical thumb movement</td>
</tr>
<tr>
<td>2</td>
<td>Push button</td>
<td>Pushing movement of hand towards screen or downwards</td>
</tr>
<tr>
<td>3</td>
<td>Swipe</td>
<td>Moving hand, wrist, or arm in vertical position</td>
</tr>
<tr>
<td>4</td>
<td>Zoom in</td>
<td>Spreading or closing thumb and one other finger in a diagonal orientation</td>
</tr>
<tr>
<td>5</td>
<td>Wave</td>
<td>Horizontal arm movement with open hand</td>
</tr>
<tr>
<td>6</td>
<td>Show hand</td>
<td>Opened had with palms facing screen</td>
</tr>
<tr>
<td>7</td>
<td>Close hand</td>
<td>Hand first opened, then closed</td>
</tr>
<tr>
<td>8</td>
<td>Previous</td>
<td>Upwards-curved vertical movement</td>
</tr>
<tr>
<td>9</td>
<td>Two</td>
<td>Stretching out index finger and middle finger, while bending ring and little finger</td>
</tr>
</tbody>
</table>

Revised criteria list

Table 10. Revised criteria list experiment 2. Underlined sentences were revised or added.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Judgment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Rotary knob</td>
<td>Rotation wrist</td>
</tr>
<tr>
<td>1</td>
<td>Remote</td>
<td>Palm directed to left or right, combined with a vertical thumb movement or: a pushing movement directed downwards</td>
</tr>
<tr>
<td>2</td>
<td>Push button</td>
<td>Pushing movement of hand towards screen or downwards</td>
</tr>
<tr>
<td>3</td>
<td>Swipe</td>
<td>Moving hand, wrist, or arm in vertical position</td>
</tr>
<tr>
<td>4</td>
<td>Zoom in</td>
<td>Spreading or closing two fingers in a diagonal orientation</td>
</tr>
<tr>
<td>5</td>
<td>Wave</td>
<td>Horizontal arm movement with open hand</td>
</tr>
<tr>
<td>6</td>
<td>Show hand</td>
<td>Opened had with palms facing screen</td>
</tr>
<tr>
<td>7</td>
<td>Close hand</td>
<td>Hand first opened, then closed or: moving a hand downwards, with fingers pointing towards the screen</td>
</tr>
<tr>
<td>8</td>
<td>Previous</td>
<td>Upwards-curved vertical movement</td>
</tr>
<tr>
<td>9</td>
<td>Two</td>
<td>Stretching out index finger and middle finger, while bending ring and little finger</td>
</tr>
</tbody>
</table>
Appendix L. Stimuli characteristics

Timeline
Each stimulus starts and ends with two seconds of the static image in which both task name and video are visible. Each ‘arm video’ starts with the same hand posture (see Figure 58). The total length of the stimulus depends on the duration of the video, but ranges between five to six seconds (see appendix J for the duration per gesture). In the combination teaching-method in which both videos are visible the total duration of the stimulus depends on the stimulus with the longest duration.

Layout
The video stimuli were scaled to make sure stimuli in the pantomime, imitation, and combination condition were equally large. In addition, all videos were positioned in the centre of the screen. See Figure 56 and Figure 57 for the exact dimensions and positions.

Figure 56. Layout of stimulus video. Dark gray area indicates position of video or animation in the imitation or pantomime condition. Videos are centered. Positions are indicated in pixels.
Figure 57. Layout of stimulus videos. Dark gray areas indicates position of video (left) and animation (right) in the combination condition. Videos are centered, and videos are separated by eight pixels. Positions are indicated in pixels.

Figure 58. Initial and final hand posture of each gesture.
Appendix M. Set-up experiment

The images below show the set-up of experiment two. The cross in the top view depicts the location at which participants were instructed to stand during the recall tasks.

Figure 59. Set-up experiment 2. The first image shows the top-view. The cross depicts the location at which participants were instructed to stand during the recall tasks. The second image shows a right-view of the set-up.
Appendix N. Questionnaire & distraction task

Dear participant,

Please answer the next five questions before continuing with the instructions:

1. What is your gender? □ male □ female

2. What is your age? ______________

3. Have you learned to communicate by using sign language (Dutch: gebarentaal)?
   □ yes □ no

4. Which hand is your dominant hand? Are you
   □ left-handed □ right-handed □ both-handed

5. What is the highest level of education (not necessarily finished)?
   □ high school □ Associate degree □ Bachelor’s degree □ Master’s degree □
   Else:______

   for Dutch participants:
   □ Middelbare school □ MBO □ HBO □ WO (Bachelor) □ WO (Master) □
   anders:______

Instructions

This second part of the experiment is a memory task in which you will be asked to remember series of shapes that you will see on the screen. Each series will be presented for only a few seconds. After viewing a series, please draw the shapes in the correct order on the next page. The drawing time is limited. Before the next series appears on the screen you will hear a short ‘beep’.

In total you will see four different shapes:  □ ▲ + ◆

In the beginning the series will be easy to remember, but they will become longer towards the end. In total you will view 17 series of shapes.

Try to draw all shapes you can remember!

To start the memory test press the right arrow-key.
When finished, please follow the instructions on the screen.
Appendix O. General instruction & informed consent

Description of experiment

The first part of the experiment includes three memory tasks in which you will be asked to remember a sequence of gestures. You will view a list of 10 videos of which each is paired with a task name. During the recall procedure you will be asked to make the hand and/or arm movements that were paired with each task name.

The second part of the experiment includes a different type of memory task. You are being filmed during both parts of the experiment. The video recordings will be used only for scientific purposes.

The whole experiment will last approximately 30 minutes. Participation in this experiment is strictly voluntary. You may choose to stop participation and withdraw from this study at any time, without providing a reason. Please indicate the researcher if you wish to stop.

Informed consent

Title of experiment: Gesture experiment 1

Responsible researcher: Eline Jansen

To be filled in by the participant:

I acknowledge that I have read this consent form and I understand its content. I understand that all the data resulting from this experiment will be handled confidentially, and my name will not be included, or in any other way associated, with the data collected in this study. All (video) data is used only for analysis and scientific purposes related to the current study.

I voluntarily participate in this research study. I acknowledge that I may stop participation at any time, without providing any reason.

Name participant: ………………………………………………………………………

Date: …………………………… Signature participant: ………………………
Appendix P. Instructions experiment

In total, participants were given four different instructions on screen; one instruction to prepare for the video stimulus presentation, one instruction to prepare for the recall part, one to instruct participants to read the instructions on the paper, and one to prepare for the fourth test. The order of the instructions is illustrated by the next image. The instructions can be found in the tables that follow.

![Diagram showing the order of instructions per stage in the experiment.](image)

*Figure 60. Overview of instructions per stage in the experiment.*
**Viewing instructions**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
</table>
| **1 All** | In this experiment you will receive instructions on the screen. Please read below the instructions for the first part of the experiment very carefully. If anything is not clear please ask the experimenter. For this part of the experiment it is important to stand at the location marked with a cross on the floor.  

This first part of the experiment is a memory task. You will view a sequence of 10 videos. |

<table>
<thead>
<tr>
<th><strong>2 Method</strong></th>
<th>imitation</th>
<th>pantomime</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each video shows an arm making a gesture and shows a task name. You should look at the gesture as if you are looking in a mirror. So, to produce the gesture you have to use your right hand, and if the arm on the screen moves to the left your arm should move to the left as well. For all gestures you need to use only one hand (right hand).</td>
<td>Each video shows a task name, and a moving interface object for which you have to think of a suitable gesture. These gestures you have to make during a subsequent recall task, so please think of the suitable gesture when viewing the videos. For all gestures you need to use only one hand (right hand).</td>
<td>Each video shows a task name, a moving interface object for which you have to think of a suitable gesture, and an arm making this gesture. So, both the interface object and the arm communicate the same gesture. These gestures you have to make during a subsequent recall task, so try to remember them. For all gestures you need to use only one hand (right hand).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>3 Practice</strong></th>
<th>practice</th>
<th>no practice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>When viewing the movies you should practice the gestures.</strong> So, try to produce the gesture when you see the 10 videos. You have to practice these gestures at shoulder height.</td>
<td>When viewing the movies you should keep your arm next to your body. <strong>You are not allowed to practice the gestures while viewing the 10 videos.</strong></td>
<td></td>
</tr>
</tbody>
</table>

| **4 All** | Your task is to remember which gesture belongs to which task name. After viewing all 10 videos, the recall procedure starts. You will be presented with the task names in random order and you need to perform the gestures you think matches the names. Please perform the gestures at shoulder height. When this is finished the whole procedure will be repeated two more times.  

So, try to remember as many gestures as possible! |
**Recall instructions**

| Condition | All | Next, you have to perform the correct gesture for each task you've just seen. Press and hold the spacebar until you remember the correct gesture you have to perform. When you have finished a gesture, press and hold the spacebar again for the next task to appear. When you do not remember the gesture anymore, release the spacebar shortly, and press and hold it again to view the next task. |

**Distraction task**

| Condition | All | Next to the computer you will find a paper with the instructions for the next task. Please sit at the desk for the next part of the experiment. Do not press to continue, before you've read and understood the instructions. |

**‘Memorability’ recall task**

| Condition | All | The shape recall task is now completed. Please stand at the location marked with a cross on the floor. Next, you will be presented with the gesture task names once more. Please, perform the correct gesture for each task. |
Appendix Q. Logistic regression recall models

A logistic regression was performed for the recall results for both learnability (based on the results of test 1, 2, and 3), and memorability (based on the results of test 4). The general models are shown in respectively Table 11 and Table 12.

Table 11. General models for recall learnability

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald</td>
<td>exp b</td>
<td>Wald</td>
<td>exp b</td>
</tr>
<tr>
<td>constant</td>
<td>8.042</td>
<td>0.842</td>
<td>8.042</td>
</tr>
<tr>
<td>method (ref: both)</td>
<td>13.399 **</td>
<td>12.664 **</td>
<td>9.258 *</td>
</tr>
<tr>
<td>imitation</td>
<td>0.315</td>
<td>0.191</td>
<td>0.585</td>
</tr>
<tr>
<td>pantomime</td>
<td>11.657</td>
<td>10.742</td>
<td>8.794 **</td>
</tr>
<tr>
<td>intuitive</td>
<td>360.081</td>
<td>390.920</td>
<td>59.329 **</td>
</tr>
<tr>
<td>practice</td>
<td>20.968</td>
<td>20.833</td>
<td>.0396</td>
</tr>
<tr>
<td>gesture type (emblem)</td>
<td>6.441 *</td>
<td>7.466</td>
<td>0.699</td>
</tr>
<tr>
<td>test 1</td>
<td>48.648</td>
<td>2.778 **</td>
<td>2.865 **</td>
</tr>
<tr>
<td>imitation*practice</td>
<td>0.278</td>
<td>1.140</td>
<td>0.278</td>
</tr>
<tr>
<td>pantomime*practice</td>
<td>9.788</td>
<td>2.181 **</td>
<td>2.181 **</td>
</tr>
<tr>
<td>gesture type*method</td>
<td>10.905</td>
<td>10.905 **</td>
<td>10.905 **</td>
</tr>
<tr>
<td>emblem*imitation</td>
<td>11.159</td>
<td>6.426 **</td>
<td>6.426 **</td>
</tr>
<tr>
<td>emblem*pantomime</td>
<td>3.724</td>
<td>3.724 **</td>
<td>3.724 **</td>
</tr>
<tr>
<td>method*intuitive</td>
<td>3.724</td>
<td>3.724 p=.05</td>
<td>3.724 p=.05</td>
</tr>
<tr>
<td>Nagelkerke pseudo R²</td>
<td>.228</td>
<td>.373</td>
<td>.386</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>460.100 (df=5)</td>
<td>800.321 (df=7)</td>
<td>832.245 (df=14)</td>
</tr>
</tbody>
</table>

* significant (0.01<p<0.05) ** significant (p<0.01)
### Table 12. General models for recall memorability

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wald</td>
<td>exp b</td>
</tr>
<tr>
<td>Constant</td>
<td>26.199</td>
<td>4.052 **</td>
</tr>
<tr>
<td>Method (ref: both)</td>
<td>17.496</td>
<td>**</td>
</tr>
<tr>
<td>imitation</td>
<td>4.290</td>
<td>2.004 ^1</td>
</tr>
<tr>
<td>pantomime</td>
<td>17.375</td>
<td>3.717 ^1 **</td>
</tr>
<tr>
<td>Intuitive</td>
<td>41.868</td>
<td>21.478 **</td>
</tr>
<tr>
<td>Practice</td>
<td>3.525</td>
<td>1.636</td>
</tr>
<tr>
<td>Gesture type (emblem)</td>
<td>3.037</td>
<td>1.562</td>
</tr>
<tr>
<td>Method*Practice</td>
<td>0.250</td>
<td></td>
</tr>
<tr>
<td>imitation*Practice</td>
<td>0.001</td>
<td>1.022</td>
</tr>
<tr>
<td>pantomime*Practice</td>
<td>0.177</td>
<td>1.320</td>
</tr>
<tr>
<td>Gesture type*Method</td>
<td>1.027</td>
<td></td>
</tr>
<tr>
<td>Emblem*Imitation</td>
<td>.235</td>
<td>1.383 ^1</td>
</tr>
<tr>
<td>Emblem*pantomime</td>
<td>0.221</td>
<td>1.350</td>
</tr>
<tr>
<td>Method*Intuitive</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>imitation*Intuitive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pantomime*Intuitive</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intuitive*Emblem</td>
<td>0.743</td>
<td>2.766</td>
</tr>
</tbody>
</table>

Nagelkerke pseudo $R^2$ | .285 | .295 |

$\chi^2$ | 109.529 (df=5) | 113.980 (df=12) |

* significant (0.01 < p < 0.05) ** significant (p < 0.01)
# Appendix R. Project proposal exam committee

## General data

<table>
<thead>
<tr>
<th>Title research proposal:</th>
<th>Teaching users how to interact with gesture based interfaces; a comparison of teaching methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Name, initials, prefixes:</td>
<td>Jansen, E.J. (Eline)</td>
</tr>
<tr>
<td>Identity number:</td>
<td>0580766</td>
</tr>
<tr>
<td>Supervisors* Name 1st supervisors:</td>
<td>Raymond Cuijpers &amp; Wijnand IJsselsteijn</td>
</tr>
<tr>
<td>Name 2nd supervisor:</td>
<td>Maurice van Beurden</td>
</tr>
<tr>
<td>Faculty/university:</td>
<td>IEIS, TU/e</td>
</tr>
<tr>
<td>Expertise provided:</td>
<td>HTI</td>
</tr>
<tr>
<td>Name 3rd supervisor:</td>
<td>John de Vet</td>
</tr>
<tr>
<td>Company/institute:</td>
<td>BG TV Innovation, Philips Consumer Lifestyle</td>
</tr>
</tbody>
</table>

* The first supervisor must be an authorised staff member from the IE&IS faculty. The second supervisor must be an authorised lecturer from TU/e or any other university, preferably from another capacity group. Specify the type of expertise provided, i.e. what aspect of the thesis he/she will supervise.

<table>
<thead>
<tr>
<th>Company/institute (if applicable) Name:</th>
<th>BG TV Innovation, Philips Consumer Lifestyle</th>
</tr>
</thead>
<tbody>
<tr>
<td>City:</td>
<td>Eindhoven</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time period Start date:</th>
<th>October 17th, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected end date:</td>
<td>April 17th, 2012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposal version Filing date of first version:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filing date of this version:</td>
</tr>
<tr>
<td>Version number student:</td>
</tr>
<tr>
<td>Version number examboard:</td>
</tr>
</tbody>
</table>
Gesture-based interfaces require users to produce body movements in a distant location. In contrast to standard interfaces the interaction-possibilities (the gestures) are not elicited by object/interface affordances (because the manipulation is indirect), but the user need to learn the gestures in advance. There are multiple methods to do this. A currently used method to teach gestures is to show a (animated) human figure performing the movements (e.g. see the Xbox Kinect). This method however, does not make use of any object affordances. A study by Wagner-Cook and Tanenhaus (2009) shows that the interaction with objects can elicit gestures. This is in line with the dominant view about the origin of gestures, saying that naturally occurring gestures are an outgrowth of mental action simulation (amongst others Hostetter & Alibali, 2008). Based on this contradiction between the use of affordances in standard interfaces and in current implementations of gesture based interfaces the following research questions were formulated.

Main research questions:

1. Which method, the observation of a human movement or an object, supports learnability and memorability of interaction gestures most?
2. What is the effect of intuitiveness of gestures on the effectiveness of the teaching methods in terms of both learnability and memorability?
3. What is the effect of practice on the memory of gesture-sets taught by both methods?

Definitions:

- **Learnability**: the number of observations that are required before participants are able to perform the correct arm movements in subsequent recall tasks.
- **Memorability**: the percentage of correct recalls after the distraction task.

Hypotheses:

1. It is hypothesized that viewing objects (pantomime) results in better recall for gestures (i.e. both learnability and memorability) compared to viewing human arms (imitation). Furthermore, a combination of both methods is expected to result in a similar performance as using pantomimes.
2. It is hypothesized that meaningful gestures are better recalled in all conditions, and have a shorter response time at recall.
3. It is hypothesized that practice has a positive effect on the recall ability and response time for gestures learned in the imitation condition. In the pantomime condition the effect of practice is expected to be less than for the imitation condition.
**Scientific and TIW relevance (max. 150 words)**

Explain that the graduation project is scientifically relevant and that it is relevant to TIW.

There are theories concerning the origin of gestures, which state that natural communicative gestures are an outgrowth of simulated action. In addition, there are studies on the variability of the movements participants make when making a gesture after viewing a human arm making a pantomime gesture, or when viewing an object. No studies are known to investigate the effect of stimulus type (showing an arm or object) on the learnability and memorability of gestures. Such a study can elucidate the processes which underlie the storage of gestures. The relevance for the company is that the results can be used as input for a gesture-based interface design.

**Method (max. 200 words)**

The hypotheses were tested in an experiment in which 120 subjects participated. The experiment was a 3 teaching-method (imitation/pantomime/both) x 2 intuitiveness (intuitive/not intuitive) x 2 practice (practice/no practice) between-subjects design. In the imitation condition participants viewed a human arm making a certain gesture. In the pantomime condition participants viewed an (custom-made) interface object that was tested to elicit a similar gesture. The correct recall of the gestures and the response time were dependent variables. Gesture type was a control variable. The stimuli in the main experiment were based on three preliminary tests. In the first pre-test (n=15) it was investigated which gestures participants would produce for a certain task in a certain context. In addition, it was tested what type of association participants had with the movement. The second pre-test (n=10) tested custom-made interface objects for the gestures that were elicited. A third test (n=8) was conducted as an iteration of the second test.

**References**

