Haptic feedback in freehand gesture interaction

Joni Karvinen

University of Tampere
School of Information Sciences
Computer Science / Int. Technology
M.Sc. Thesis
Supervisors: Roope Raisamo and Jussi Rantala
January 2015

University of Tampere
School of Information Sciences
Computer Science / Interactive Technology
Joni Karvinen
M.Sc. Thesis, 66 pages
January 2015

In this thesis work, haptic feedback in gesture interaction was studied. More precisely, focus was on vibrotactile feedback and freehand gestural input methods. Vibrotactile feedback methods have been studied extensively in the fields of touch-based interaction, remote control and mid-air gestural input, and mostly positive effects on user performance have been found. An experiment was conducted in order to investigate if vibrotactile feedback has an impact on user performance in a simple data entry task. In the study, two gestural input methods were compared and the effects of visual and vibrotactile feedback added to each method were examined. Statistically significant differences in task performance between input methods were found. Results also showed that less keystrokes per character were required with visual feedback. No other significant differences were found between the types of feedback. However, preference for vibrotactile feedback was observed. The findings indicate that the careful design of an input method primarily has an impact on user performance and the feedback method can enhance this performance in diverse ways.

Key words and terms: vibrotactile feedback, freehand gesture interface, Leap Motion

Contents

1 Introduction

The act of gesturing is a natural way of communication for humans and it is considered to be more expressive than speech alone. Gestures can be used to convey a variety of meaningful information. They can be used to express an idea, depict a part of an utterance, to convey culturally specific meanings or simply point to objects in space. They are interpreted in the current situation and this interpretation varies between individuals in different social and cultural environments. For these reasons gesticulation is a powerful medium in the communication between individuals.

For the very same reasons, gesture interaction has not yet been utilized in its full potential in interaction between a human and computer. Challenges are faced in the development of recognition techniques as well as in interaction design due to the ambiguous and multifaceted nature of gestures. Gesture interaction has been an area of extensive research and solutions have been suggested to most problems.

However, one major shortcoming in freehand gesture interfaces has been the lack of haptic feedback. Users have had to rely mainly on visual, aural or proprioceptive feedback. Considering the versatility of the sense of touch, useful information is lost during the interaction. Meaningful information can be conveyed via sense of touch, Braille writing system being an example, and tactile sensations can also be emotionally charged. Solving the problem of absence of haptic feedback is difficult especially with contactless interfaces because feedback must be completely artificial. However, in the recent year clever contactless solutions for generating tactile feedback have already been proposed.

The main goal of this thesis is to investigate how vibrotactile feedback could enhance user performance in gestural interaction. Another goal is to find out what kind of possibilities gestural input might have and what sort of problems could be faced in the design of novel interaction methods.

The thesis consists of three parts. Chapter 2 focuses on gestural input. The chapter begins with a definition of gesture and an overview of gesture classifications. After that, a unified outlook on the issue of naturalness is constructed by combining different perspectives on the subject. Through a short presentation of a few application domains, I proceed to talk about the design of gestural interaction. The design section scrutinizes up-to-date design heuristics, properties of a desirable gesture command, ergonomic factors as well as techniques to enhance learnability of the interaction. After the design issues have been covered, experimental results from the studies comparing freehand gesture interaction to remote control and touch-based input are presented. A couple of software tools for interaction design are presented at the end of the chapter.

Chapter 3 deals with haptic feedback and it comprises four subsections. First, the term haptics is defined. Second, four-channel model of mechanoreception and functions of each skin receptor are explained. Third, in order to find basis for the fuzz around the field of haptics, evidence for and against tactile feedback is offered. The impact on user performance when touch is acting as a single modality and as a part of multimodal feedback are investigated. User preferences have also been taken into account. Whenever appropriate, the knowledge gained from the experiments is applied to the design of gestural interfaces. The chapter ends with an overview of tactile technologies.

In addition to literature review, a major part of the thesis work was the implementation of an application and conducting a user experiment to evaluate its use. The application is a virtual numeric pad that is controlled with Leap Motion using two distinct input methods. Visual and vibrotactile feedback styles associated to each input method were also created. In the experiment these input methods were compared and the effects of feedback types on user performance were studied. The experiment and the application are described in Chapter 4.

Finally, topics covered in this thesis are discussed in a wider perspective in Chapter 5. The direction of the development of gestural interfaces and haptic feedback in the future is speculated. Discussion is also expanded on the topics that were purposely left out from closer examination in the Chapters 2 and 3. In addition, the results of the experiment are summarized.

2 Gestures

At least two forms of gesture interaction can be distinguished. Nancel et al. [2011] discriminate between the terms *freehand* and *mid-air* gesture interaction. Freehand techniques are based on motion tracking whereas mid-air techniques require the user to hold an input device. In this work, interest is on freehand method and therefore this chapter discusses the design and implementation of freehand interaction.

In the first section, definitions of gesture are provided. The second section presents established gesture classifications as well as categorizations specifically tailored for human-computer interaction. Naturalness of gesture interaction is contemplated in the third section. Before going deeper into the design issues, a few application domains in which mid-air gesturing has been found to be appropriate are presented. The design section proceeds from the design of gesture commands to the design of interaction. Heuristics for design, properties of a meaningful and comfortable gesture and the learnability of gesture interaction as a whole are discussed. After this, results from the studies comparing mid-air techniques to other interaction methods are offered and reflected upon. In the last section, two interaction design tools are shortly presented.

2.1 *What are gestures?*

According to Oxford Dictionary of English (2010, 3rd ed.) gesture is ”a movement of part of the body, especially a hand or the head, to express an idea or meaning”. Quek et al. [2002] expand upon this definition and consider facial expressions and gaze shifts also as gestures. For Mitra and Acharya [2007] gestures have two intentions. One is to convey meaningful information which can be dependent on the spatial, pathic, symbolic and affective information. The other is to interact with the environment.

For McNeill [1992, 2006] the definition of gesture is equal to that of gesticulation. In his view gesture and language are integrated and should be viewed as a single system. Speech and gestures are used to complement each other.

Kendon's [2004] view slightly differs from McNeill's. According to his definition gesture is ”a name for visible action when it is used as an utterance or as part of an utterance” [Kendon, 2004, p. 7]. Thus, for Kendon gesture itself can be a linguistic expression such as an emblem or a sign in sign language. However, not any visible bodily action is regarded as gesture. Kendon specifies that gesture is ”a label for actions that have the features of manifest deliberate expressiveness” [Kendon, 2004, p. 15]. Involuntary or habitual movements are not referred to as gestures but what's essential is the communicative intent of an actor. Nonetheless, the actual meaning of a certain gesture is subject to social convention and cultural context. Gestures and their meanings vary between individuals and may even be different for the same person in different situations [Mitra and Acharya, 2007].

Furthermore, a distinction is made between postures and gestures. Hand postures are seen as static finger configurations without hand movement and hand gestures as dynamic movements which may or may not involve finger motion [Mitra and Acharya, 2007].

2.2 *Classification of gestures*

In this section several gesture classifications are presented. An overview of the four already established classifications is provided and later in this section, categorizations tailored for human-computer interaction are presented. Most of the classifications bear resemblance to each other but emphasize different aspects of gestural communication. I intend to point out these similarities and find connections between categorizations.

2.2.1 Gesture classifications for human-human interaction

Four classifications of Efron, Cadoz, Kendon and McNeill are presented. Gesture taxonomies presented here have been created in the fields such as linguistics and anthropology. Thus, they provide a universal perspective on communicative properties of gesturing. Later, the attention is drawn on the classifications which have been made in the area of human-computer interaction (HCI).

2.2.1.1 Efron

Efron's [1941] classification is one of the earliest attempts to categorize discursive human gestures. His work has influenced many subsequent taxonomies such as the works of Kendon [1988] and McNeill [1992] which are later introduced in this chapter.

Efron distinguishes two main types of gesture: *logical or discursive gestures* and *objective gestures*. Gestures which do not portray any object of reference but thought process related to speech are called logical. These gestures do not depict what the speaker is talking about but instead refer to the elements of speech itself. Two sub-categories of logical gestures are *batons* and *ideographics*. Batons are rhytmic gestures which are used to highlight certain words or phrases in an utterance. Ideographic gestures are performed to present the path or direction of a thought pattern.

Objective gestures, on the contrary, convey meaning indenpendently of speech and they can be further divided into *deictic*, *physiographic* and *symbolic or emblematic* gestures. Deictic gestures are also called pointing gestures. Physiographic gestures can be split into two subcategories: *iconographics* which depict the form of an object or spatial relationships, and *kinetographics* that depict a bodily action. Symbolic, or emblematic gestures, are conventionalized and culturally specific signs that are not governed by any formal grammar [Quek et al., 2002; McNeill, 2006]. ”Thumbs up” or ”Ok” (thumb and forefinger joined together) signs are examples of such gestures.

2.2.1.2 *Cadoz*

Cadoz [1994] classifies hand movements into three groups according to their function: *ergotic*, *epistemic* and *semiotic* gestures. Ergotic gestures are used for manipulating the physical world such as interacting with a touchscreen. Epistemic gestures are performed to explore the environment through haptic sensing and proprioception. For example, checking the presence of a wallet in the back pocket. Semiotic gestures are used to communicate meaningful information in human-human interaction. Semiotic hand movements can be further extended by McNeill's classification and according to their linguisticity by Kendon's continuum [Mulder, 1996].

2.2.1.3 *Kendon's continuum*

Kendon [1988] arranges gestures along a continuum (depicted in Fig. 1). Moving from left to right in Fig. 1 the necessity of accompanying speech decreases and the degree to which gesture has properties of language increases [McNeill, 2006].

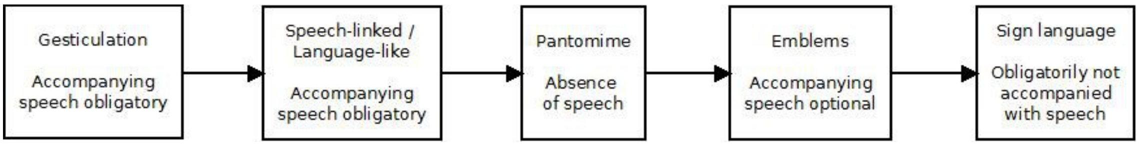


Fig. 1. Kendon's continuum.

Gesticulation is also referred as coverbal gestures which describes the concept accurately. The act of gesticulating is characterized as depictive or iconic free-form gesturing, which is not taught and typically accompanying speech but also other modalities can be involved [Quek et al., 2002; Karam and Schraefel, 2005].

Language-like gestures are similar to gesticulation. However, gesticulation is mostly performed synchronously with coexpressive speech whereas language-like gestures are part of the sentence itself [McNeill, 2006]. They are used to fill a linguistic gap and complete the sentence structure (”The bird flew like [a gesture depicting flapping wings]”).

Pantomimes are iconic gestures or sequences of such gestures which convey a narrative line [Quek et al., 2002; McNeill, 2006]. Pantomimes are produced without speech.

Kendon referred to *emblems* also as quotable gestures. They can occur with speech but are also meaningful on their own.

Sign language is linguistically based and it is characterized by grammatical and lexical specification [Quek et al., 2002]. Sign language is not necessarily accompanied by speech since simultaneous speaking and signing may interfere both [McNeill, 2006].

2.2.1.4 McNeill

McNeill's [1992] classification expands gesticulation and language-like categories on Kendon's continuum (Fig. 1). Moreover, McNeill's work is largely based on Efron's classification. Gestures are divided into four categories: *iconic*, *metaphoric*, *deictic* and *beat*. Any of these gestures can be cohesive which means that they are used to tie together parts of the discourse which are semantically related but temporally separated. Furthermore, gestures are grouped to imagistic or non-imagistic types depending on whether they depict imagery.

Iconic gesture is one that bears ”a close formal relationship to the semantic content of speech” [McNeill, 1992, p. 78]. Efron used the term physiographics describing similar gestures. Metaphoric gestures differ from iconic ones in that they present an image of an abstract concept whereas iconic gestures refer to a concrete event or an object [McNeill, 1992]. In Efron's classification, metaphoric gestures were referred to as ideographics. Iconic and metaphoric gestures both belong to imagistic type.

Deictic gesture is a pointing movement which is usually performed with the pointing finger but also any extensible object or body part can be used [McNeill, 1992]. Beats do not convey meaning but are used to express the structure and rhythm of speech or stress specific words and phrases. Efron referred to gestures of this kind as batons. According to McNeill's [1992] definition beats are rapid flicks of the fingers or hand that have two movement phases – in/out, up/down etc. and can be performed in the periphery of the gesture space (the lap, an armrest of the chair, etc.).

In McNeill's classification the relationship between narrative and gesturing is a fundamental basis. The key idea is that speech and gestures are coexpressive, convey information about the same scenes, the same ”idea units”, and each can include what other leaves out [McNeill, 1992; Quek et al., 2002].

2.2.2 Gesture classifications for HCI

Classifications suggested by Efron, Cadoz, Kendon and McNeill all describe discursive gestures in human-human communication. Therefore, these categorizations are not directly applicable to human-computer interaction. In this section three categorizations that are especially tailored for HCI are presented.

2.2.2.1 Taxonomy of Karam and Schraefel

Comprehensive taxonomy of Karam and Schraefel is based on a literature review and a framework of Quek et al. [2002] (semaphores, manipulation, gesture-speech approaches). In their unique approach they categorize gestures in terms of four key elements: gesture styles, gesture enabling technology, application domain and system response.

Karam and Schraefel [2005] divide gesture styles into five categories: deictic, manipulative, semaphoric, gesticulation and language gestures. According to Karam and Schraefel [2005, p. 4] *deictic* gestures ”involve pointing to establish the identity or spatial location of an object within the context of the application domain.” In their definition of *manipulative* gestures they refer to that proposed by Quek et al. [2002]. Manipulative gestures are ”those whose intended purpose is to control some entity by applying a tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated” [Quek et al., 2002, p. 172]. However, direct manipulation such as dragging, moving or clicking objects are not considered as gestures because the system must be able to interpret the actions of the user and translate the gesturing as a command until it can be categorized as manipulative [Karam and Schraefel, 2005]. This definition is what makes manipulative gestures different from deictic gestures. Quek et al. [2002] also point out that the dynamics of hand movement in manipulative gestures differ significantly from conversational gestures and they may be aided with visual, tactile or force feedback from the object being manipulated. For instance, pressure can be used as additional information on the table-top surfaces.

Again, borrowing the definition provided by Quek et al. [2002, p. 172], *semaphoric* gestures are ”any gesturing system that employs a stylized dictionary of static or dynamic hand or arm gestures”. Semaphoric gestures differ from manipulative gestures in that they are considered to be communicative and do not typically require feedback control for manipulation [Quek et al., 2002]. Efron [1941] and Kendon [1988] referred to these kinds of gestures as emblems. Strokes or other similar gestures are also considered semaphoric. A gesture can be either a static pose or dynamic whenever movement is involved. Semaphoric hand use covers only a small portion of the typical gesturing in communication because expressions are learned and consciously used, thus considered not natural and providing little functional utility [Quek et al., 2002]. *Gesticulation* refers to the similar concept as in Kendon's and McNeill's models. Furthermore, like Kendon, Karam and Schraefel also include *language* gestures (finger spelling, sign language) as a distinct category.

Karam and Schraefel split gesture enabling input technologies into two classes: non-perceptual and perceptual. Non-perceptual input involves technologies that require physical contact with the device or object that is used to perform the gesture whereas perceptual input does not.

As one key element in their taxonomy, Karam and Schraefel present the classification of gesture focusing on application domains they are applied to. These include virtual/augmented reality, desktop/tablet PC applications, CSCW (computer-supported cooperative work), 3D displays, ubiquitous computing and smart environments, games, pervasive and mobile interfaces, telematics, adaptive technology, communication interfaces and gesture toolkits.

As a final categorization element, Karam and Schraefel suggest different output technologies. They separate these technologies into three categories: audio, visual (2D and 3D) and CPU command responses.

2.2.2.2 Taxonomy of surface gestures

The taxonomy proposed by Wobbrock et al. [2009] is based on their elicitation study in the context of surface computing (see Table 1). They classify gestures along four dimensions which are form, nature, binding and flow. Each of these dimensions include multiple categories.

Form dimension involves a pose of the hand, either *static* or *dynamic*, and a path along which the hand possibly moves. *Nature* dimension is further divided into symbolic, physical, metaphorical and abstract gestures. *Symbolic* gestures are visual depictions, comparable to what Kendon referred as emblems or semaphoric gestures in Karam and Schraefel's taxonomy. *Physical* gestures are used to manipulate objects on a screen. *Metaphorical* gestures represent action or depict the form of the referent. When the connection between the gesture and the referent is arbitrary, the gesture is considered *abstract*.

The *binding* dimension defines what information is required about the location where the gesture is being performed. *Object-centric* means that the gesture affects only the object on which it is being performed. *World-dependent* gestures are performed on a specific location on the screen whereas *world-independent* gestures can occur anywhere on the display. *Mixed dependencies*, for instance, can occur for two-handed gestures where one hand is required to act on an object and the other can act anywhere on the screen. The gesture's *flow* can be either *discrete* which means that response occurs after completion of the gesture, or *continuous* which means that response occurs while the user acts such as during resizing of an object.

The taxonomy of Wobbrock and others can be applied to two-dimensional surface interaction. Ruiz et al. [2011], for their part, focus on three-dimensional interaction and propose a taxonomy of motion gestures in mobile interaction context.

TAXONOMY OF SURFACE GESTURES		
Form	*static pose*	Hand pose is held in one location.
	dynamic pose	Hand pose changes in one location.
	static pose and path	Hand pose is held as hand moves.
	dynamic pose and path	Hand pose changes as hand moves.
	one-point touch	Static pose with one finger.
	one-point path	Static pose & path with one finger.
Nature	*symbolic*	Gesture visually depicts a symbol.
	physical	Gesture acts physically on objects.
	metaphorical	Gesture indicates a metaphor.
	abstract	Gesture-referent mapping is arbitrary.
Binding	*object-centric*	Location defined with respect to object features.
	world-dependent	Location defined with respect to world features.
	world-independent	Location can ignore world features.
	mixed dependencies	World-independent plus another.
Flow	*discrete*	Response occurs *after* the user acts.
	continuous	Response occurs *while* the user acts.

Table 1. Taxonomy of surface gestures suggested by Wobbrock et al. [2009].

2.2.2.3 *Motion gestures for 3D mobile interaction*

The taxonomy of motion gestures proposed by Ruiz et al. [2011] contains two classes of taxonomy dimensions: gesture mapping and physical characteristics. Both of these dimensions are further divided into three subdimensions. Furthermore, these additional dimensions are separated into categories. The taxonomy is presented in Table 2. Ruiz and others clarify that motion gestures refer to gestures in which a user also translates or rotates the device instead of just acting on a touchscreen.

Gesture mapping dimension describes how gestures are mapped to device commands and it includes nature, context and temporal dimensions. *Nature* defines the gesture mappings to physical objects and it is segmented into metaphorical, physical, symbolic and abstract categories. *Metaphorical* gesture is acting on a physical object other than a phone. *Physical* gesture means direct manipulation. When the user visually depicts a symbol, it is a *symbolic* gesture and when the gesture mapping is arbitrary, it belongs to an *abstract* category.

A gesture in the *context* dimension can be either an *in-context* or an *out-of-context* gesture. For example, placing the phone to the head to answer a call is an in-context gesture whereas a shaking gesture to return to the home screen is considered an out-of-context gesture.

Temporal dimension is comparable to flow dimension in the taxonomy provided by Wobbrock et al. [2009]. In a similar fashion, the gesture can be either *discrete* or *continuous* depending on whether the action occurs after or during a gesture is performed.

TAXONOMY OF MOTION GESTURES		
Gesture Mapping		
Nature	Metaphor of physical	Gesture is a metaphor of another physical object
	Physical	Gesture acts physically on an object
	Symbolic	Gesture visually depicts a symbol
	Abstract	Gesture mapping is arbitrary
Context	In-context	Gesture requires specific context
	No-context	Gesture does not require specific context
Temporal	Discrete	Action occurs after completion of gesture
	Continuous	Action occurs during gesture
Physical Characteristics		
Kinematic Impulse	Low	Gestures where the range of jerk is below 3m/s3
	Moderate	Gestures where the range of Jerk is between 3m/s3 and 6m/s3
	High	Gestures where the range of Jerk is above 6m/s3
Dimension	Single-Axis	Motion occurs around a single axis
	Tri-Axis	Motion involves either translational or rotational motion, not both.
	Six-Axis	Motion occurs around both rotational and translational axes
Complexity	Simple	Gesture consist of a single gesture
	Compound	Gesture can be decomposed into simple gestures

Table 2. Gesture taxonomy proposed by Ruiz et al. [2011].

Physical characteristics dimension includes kinematic impulse, dimension and complexity. *Kinematic impulse* is categorized as low, moderate or high depending on the range of jerk (rate of change of acceleration) applied to the phone throughout the gesture. *Dimension* describes how many degrees of freedom are involved in the movement and it can be either single-axis, tri-axis or six-axis. *Complexity* is split into two categories, *simple* and *compound* gesture. Simple gesture consists of only one gesture but compound gestures can be decomposed into simple gestures.

2.3 *Naturalness of gesture interaction*

Along with advances in technology, new interactions have been labeled as ”natural user interface” (NUI). The term includes not only vision-based techniques but also other techniques such as voice commands, pen-based input, face interfaces and multitouch gestural input. What exactly is meant by natural has received a variety of loose definitions. Some define it as the mimicry of the real world, some associate it with intuitiveness and some explain it from the usability viewpoint.

Rhetorics like Microsoft Kinect's marketing slogan ”You are the controller” promises that NUI allows a user to become the interface and there's no more requirement to learn specific techniques to operate devices. That users can act and communicate with computers through physical movements and speech as they would naturally in real life. These claims contain the idea that computers could interpret and understand the user's every intent, no matter how ambiguous or arbitrary the action, then react to it appropriately despite the context interaction takes place in and all this will be accomplished as smoothly as in human-human interaction.

Whether new interactions can be considered natural or not has been a topic of debate in the literature. These claims have been strongly criticized by Norman [2010]. His critique is targeted at new conventions which neglect well-established standards and guidelines of design. Norman states that natural user interfaces are no more natural than any other form of interaction and points out limitations of gestural input as the only choice of interaction.

In his view learnability and memorization of gesture commands are difficult due to the incompatibility of gestures and their expected effects. Gesture mappings may be natural for few simple tasks but defining gestures for abstract and complex actions leads to unnatural and arbitrary commands. According to Blackler and Hurtienne [2007] intuitive design is built upon familiar features and it utilizes the users' prior knowledge from other experiences resulting in fast and unconscious decision-making during interaction. Due to the violation of usability standards and introduction of new unfamiliar conventions intuitive use cannot be achieved.

Although Norman does not fully approve of the concept of natural interaction, he does acknowledge its advantage in expanding interaction arsenal but only if they're utilized in appropriate contexts and as an addition to other forms of interaction. Even though Norman's criticism is pertinent and summarizes the problems in NUI design, it is focused on what O'Hara et al. [2013] refer to as representational concern, that is, debating about the naturalness of the interface itself.

For O'Hara and others, just like Norman, technology itself is not natural. In their view, naturalness is always attached to social context. Essential is the concept of *community of practice.* People experience world and make it meaningful through practice. The actions people perform with technology are fitted to the social settings and practices in their particular community. The properties of technology are interpreted and made meaningful differently in different communities depending on how the system entails potential for action in their particular practices.

Environment can enable or constrain how actions are performed and how they can be fit to a particular social setting. First of all, there has to be enough space in order to use gestural systems appropriately. If the user does not have enough freedom to move, natural use can be seriously hindered. Technological limitations can also determine the user's freedom of movement. In the case of multiple users, system may encounter tracking problems if the users are too close to each other. Social environment in which the system is being used sets rules for what can be done. Norms and expectations of appropriate movements affect the way gestural interfaces are used. If a person is using a public display system, waving hands in the air or other movements of similar nature could embarrass the user. Furthermore, the appropriateness of individual gestures is perceived differently in different cultures.

Instead of focusing on interface alone or social contexts, Wigdor and Wixon [2011] turn their attention to users. For them, natural refers to the way users interact with the product and how they feel during the experience. They define natural user interface by three elements: enjoyable, leading to skilled practice and appropriate to context. An interface must have all of these elements.

One of the promises of NUI is to add fun into the interactions and make using the product feel completely comfortable. But before usage feels natural, new conventions have to be learned. Wigdor and Wixon do not associate naturalness with intuitiveness. According to them non-traditional methods have to be designed in a new way and reliance on familiar features or metaphors is not suitable. One of the goals of NUI design is to efficiently support the development of skilled behaviour in order to interaction with the system continue to feel natural and enjoyable to its users. NUI should provide comfortable user experience in a context where gestural input is appropriate and natural to most of its expert users.

In conclusion, concentrating on technology alone in defining naturalness is inadequate. Naturalness is not the same as the user becoming the interface and controlling the machine through body movements. Naturalness lies in the potential actions new technologies enable and how these can be made meaningful within certain communities and social settings. Thus, natural use varies between different user groups. Important is that using the product feels natural and creates an experience of mastery and pleasantness. Using keyboards and mouse, touch-based interaction or natural language interfaces are not natural either. However, without proper technology that enables gestural control, the goals of NUI and experience of naturalness can never be achieved.

2.4 *Application domains*

There are situations where gestural interaction is especially useful if not necessary. In this short overview, three such situations are presented.

Freehand gesture interaction has one advantage that cannot be achieved with other interface types. It allows sterile interaction which is highly important for instance in surgical environments. Wachs et al. [2008] have developed a system called Gestix, a hand gesture system for MRI manipulation in an EMR image database (Fig. 2). The system allows sterile interaction that is rapid and easy to use since surgeons are highly skilled in working with their hands. Hand gestures are recognized accurately up to five meters from the camera. Therefore, delays caused by a surgeon visiting the main control wall away from the patient's side are avoided. Interaction becomes also faster because surgeons can manipulate images from a distance on their own without needing to instruct other colleagues to browse the images.

Fig. 2. A surgeon browsing medical images with Gestix. [Image source: Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan, Vision-based hand-gesture applications. *Commun. ACM,* **54,** 2 (February 2011), 60-71.]

Outside operating rooms, gestural interaction enables control of public displays from a distance. StrikeAPose is an interactive public display game created by Walter et al. [2013] primarily for research purposes. The player's mirror image is shown on the screen and the user can use this image to play with virtual cubes which are tossed into specific targets to collect points (Fig. 3). A teapot gesture is performed to add a doctoral hat or a funny bunny mask to the user's contour. StrikeAPose has been developed as an entertainment application but similar interactive method can be utilized for control of information displays as well. Also, using gestural commands from a distance could help users who cannot easily use touch screens such as disabled people.

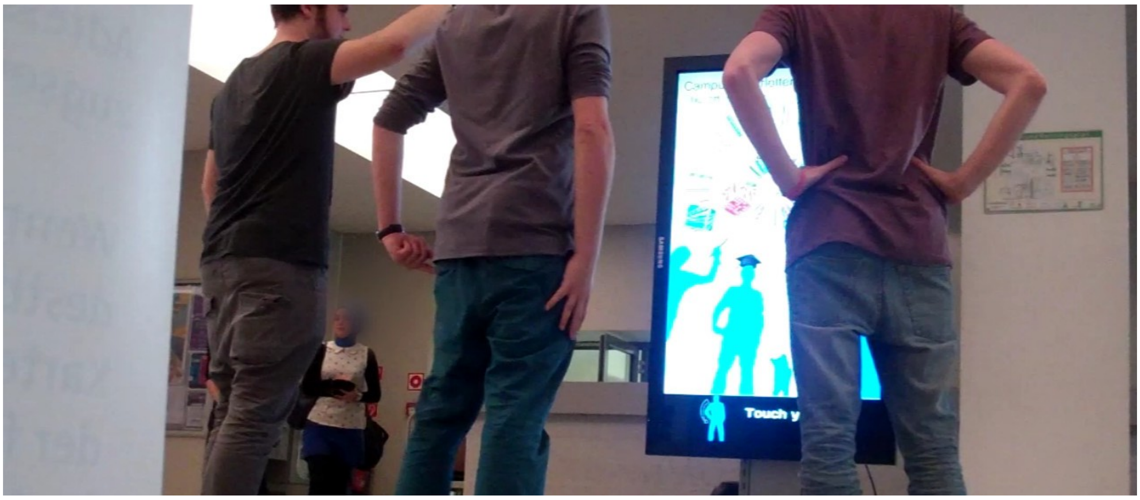


Fig. 3. Passers-by playing StrikeAPose. Contours of the players are shown on the screen. The player in the middle performs a teapot gesture and a doctoral hat is added to the mirror image. [Image source: http://www.rwalter.de/projects/strikeapose/]

Gestural interaction for interactive TV control has been a subject of extensive investigation in the literature. Hand gesturing in free air has become an appropriate choice for control as screen sizes have been constantly increasing. Gesture control has the potential to make interaction fluent and remove the need for remotes. Defining what sort of gestures should be used and how exactly these gestures ought to be performed are questions for which answers are anything but simple and straightforward. In the next section, I clarify the difficulties and seek possible solutions for the design of freehand interaction.

2.5 *Design of gesture interfaces*

This section is dedicated to the design issues of freehand gesturing. The section begins with the introduction and comparison of heuristics for freehand gesture interaction and traditional GUI interaction. After this, the focus is shifted to the properties of gesture commands and the design of gesture vocabularies. I also seek to answer how knowledge obtained from the classifications of gestures can be utilized in the design of gestures and how ergonomic factors should be taken into account. In the discussion of learnability at the end of the section I concentrate on the interaction more broadly and I bring up design issues that could aid in embracing new methods for human-computer interaction if properly implemented.

2.5.1 Heuristics

Based on their literature review Maike et al. [2014] compiled a set of 23 heuristics for the design of natural user interfaces (Table 3). Heuristics are divided into four categories: interaction, navigation, user adoption and multiple users.

Interaction category contains nine heuristics. The first two of these, *operation modes* and *”interactability”*, focus on the interface design. The system should provide different operation modes and a transition between modes should be smooth. It should also be clear to the user which objects on the screen are selectable and ”interactable”. Two heuristics take into account the technical implementation of a system. These heuristics are *responsiveness* and *accuracy* which state that tracking and detection of input gestures should be accurate and recognition should happen in real-time. Three heuristics in the list address the utilization of metaphors. These are *identity*, *metaphor coherence* and *distinction*. Metaphors have to make sense and be easily understood (identity), they should have a clear relationship with the functionalities of the interface (metaphor coherence) and they should be distinctive from one another (distinction). The last two remaining heuristics advice to design gestures that do not cause fatigue *(comfort)* and utilize gesture interfaces for the tasks they are especially good for *(device-task compatibility)*.

Navigation category consists of four heuristics. An interface should support *active exploration* in order that learning can be constructed and the user can smoothly develop skilled practice. This can be achieved through *guidance.* For Wigdor and Wixon [2011] these same ideas were essential in making the interaction feel natural. Also, as in GUI or any other type of interface design, users should know where they are at every given moment and moving from place to place without getting lost should be ensured. This is referred to as *wayfinding.* In addition, the actual *space* interaction takes place in may limit the possibilities of interaction methods and gesture commands should be designed accordingly.

User adoption category contains six heuristics. Heuristics in this category are guidelines to make the gesture interface more appealing and more efficient and easier to use than current systems. In *competition* between traditional and new alternatives, new interaction methods should beat the older ones in efficiency, ease of use and engagement. Some of these heuristics are partially related to those previously explained. Systems and devices should also compete with the price *(affordability)*. *Learnability* is related to the support for active exploration and novice-to-expert transition but this particular heuristic emphasizes that the amount of time required to learn the task should be kept to a minimum, depending on the difficulty of a task and frequency of use. *Engagement* can also enhance active exploration. *Familiarity* is close to intuitiveness. It does not necessarily mean that an interface should resemble non-NUI interfaces graphically or mimic their actions although metaphors can be borrowed from GUI interaction and refined if necessary. More important is that there is a coherence between metaphors and functionalities. *Social acceptance* states that using a gesture interface should not embarrass the user.

The last category focuses on multiple users. *Learning* is related to active exploration and learnability issues discussed earlier. The difference is that users can learn together monitoring and copying actions of one another. *Conflict* heuristic is a guideline for technical implementation. The system should be able to recognize simultaneous inputs and interpret them separately. The last two, *parallel processing* and *two-way communication*, concentrate on how tasks can be performed simultaneously. Besides group view, each user should have a personal view and the users should be able to communicate with each other while working either at a distance or in the same location. The list proposed by Maike and others is comprehensive and it covers design issues diversely but other suggestions also exist.

Table 3. 23 heuristics suggested by Maike et al. [2014].

Interaction	
Operation modes	Provide different operation modes, each with its own primary information carrier (e.g., text, hypertext, multimedia...). Also, provide an explicit way for the user to switch between modes and offer a smooth transition.
“Interactability”	Selectable and/or “interactable” objects should be explicit and allow both their temporary and permanent selection.
Accuracy	Input by the user should be accurately detected and tracked.
Responsiveness	The execution of the user input should be in real time.
Identity	Sets of interaction metaphors should make sense as whole, so that it is possible to understand what the system can and cannot interpret. When applicable, visual grouping of semantic similar commands should be made.
Metaphor coherence	Interaction metaphors should have a clear relationship with the functionalities they execute, requiring a reduced mental load.
Distinction	Interaction metaphors should not be too similar, to avoid confusion and facilitate recognition.
Comfort	The interaction should not require much effort and should not cause fatigue on the user.
Device-Task compatibility	The tasks for which the NUI device is going to be used have to be compatible with the kind of interaction it offers (e.g., using the Kinect as a mouse cursor is inadequate).

Navigation	
Guidance	There has to be a balance between exploration and guidance, to maintain a flow of interaction both to expert and novice users. Also, shortcuts should be provided for expert users.
Wayfinding	Users should be able to know where they are from a big picture perspective and from a microscopic perception.
Active Exploration	To promote the learning of a large set of interaction metaphors, a difficult task, active exploration of this set should be favored to enhance transition from novice to expert usage.
Space	The location in which the system is expected to be used must be appropriate for the kinds of interactions it requires (e.g., full body gestures require a lot of space) and for the number of simultaneous users.
User adoption	
Engagement	Provide immersion during the interaction, at the same time allowing for easy information acquiring and integration.
Competition	In comparison with the equivalent interactions from traditional non-NUI interfaces, the NUI alternative should be more efficient, more engaging and easier to use.
Affordability	The NUI device should have an affordable cost.
Familiarity	The interface should provide a sense of familiarity, which is also related to the coherence between task and device and between interaction metaphor and functionality.
Social acceptance	Using the device should not cause embarrassment to the users.
Learnability	There has to be coherence between learning time and frequency of use; if the task is performed frequently (such as in a working context), then it is acceptable to have some learning time; otherwise, the interface should be usable without learning.
Multiple Users	
Conflict	If the system supports multiple users working in the same task at the same time, then it should handle and prevent conflicting inputs.
Parallel processing	Enable personal views so that users can each work on their parallel tasks without interfering with the group view.
Two-way communication	If multiple users are working on different activities through the same interface, and are not necessarily in the same room, provide ways for both sides to communicate with each other.
Learning	When working together, users learn from each other by copying, so it is important to allow them to be aware of each other's actions and intentions.

Zamborlin et al. [2014] propose four properties gesture interfaces should provide to create interaction which is as effective as possible. The first one is *continuous control*. User movements and recognition processes should be synchronised continuosly. The system should also be prepared for the continuous changes in gestures.

The second property emphasizes the importance of building a system that is *tailorable for specific context*. Users should be able to define their personal gesture vocabularies. This way users could themselves adapt their interaction to different contexts and environments. Furthermore, users could modify gestures later as their expertise develops.

The third property is *meaningful feedback*. Users should access as much information as possible synchronously and continuously and the system should also provide information at different levels of detail. Moreover, users should not be forced to rely solely on visual feedback but they should be given a possibility to choose from a range of alternatives the most appropriate to the task at hand.

The last property is to *allow expert and non-expert use*. Defining gestures should be sufficiently simple, quick and straightforward and functionality easily accessible.

One should not forget Nielsen's [1994] heuristics which are applicable to the design of gestural interfaces even though they are targeted for the design of GUI interfaces. It can be argued that modern gesture interfaces also violate some of the traditional design instructions. Often users are forced to remember arbitrary gesture commands which do not comply with the rule of relying on recognition than recall. Also, designer-created gestures and interaction methods may not always meet with the expectations of users. Good error handling and continuous information provided to the user are virtues also in gestural interface design but maybe dialogue-based communication emphasized by Nielsen is no more meaningful or efficient for gesture interaction, at least in a manner it has been used in graphical user interfaces. Continuous information should be embedded into the interaction in order to avoid constant interruptions. The dialogue with the system could also be carried out by using gesture commands. Although some of the heuristics are not directly applicable for gesture interaction, as higher level guidelines Nielsen's heuristics are still worth following.

Similarities are apparent when the lists are put in comparison. All of the lists bring up support for novice and expert use, that users should have fluent continuous control, that interaction metaphors and functionalities should match and users ought to have access to all the necessary information at every moment.

Novel technology has raised new questions as well. A lot more emphasis is placed on the technical implementation such as accuracy or responsiveness. Besides technical issues, one has to take into account the environment in which the interface is being used. For example, full-body gesture interfaces require a lot of space. Simultaneous users has to be taken into account as well. Sociality in a wider context is also addressed.

Social acceptance is an important part of user adoption. Using an interface should not cause embarrassement to the user. Inappropriate gesture commands or the user not knowing how to operate the system after a short amount of time might be reasons for users to abandon novel technology.

2.5.2 What kind of gestures should be designed?

Earlier in this chapter several gesture classifications were presented. In this section I provide study results that clarify which ones of the gesture categories are most likely to be preferred by the users. Other properties of gesture commands are also examined. Findings presented here are largely based on elicitation studies. Elicitation approach is a widely used method for constructing gesture languages. In elicitation studies users are asked to come up with gestures of their choice for certain actions or users are being studied in a natural environment.

Elicitation studies have revealed a great deal of interesting findings about user preferences for freehand gestures. It appears that one-handed, simple gestures are preferred over two-handed, complex gestures [Wu and Wang, 2013; Vatavu, 2012]. However, opposite preferences were found in the study of Nancel et al. [2011]. Two-handed techniques resulted in faster movement times compared to one-handed techniques. Perhaps contradictory preferences can be explained by the context. In the studies of Wu and Wang [2013] and Vatavu [2012] the goal was to come up with gesture commands for basic TV controls whereas Nancel and others studied gestures in mid-air pan and zoom tasks for wall-sized displays. Separating two actions, specifying the focus of expansion with a dominant hand and controlling zooming and panning with a non-dominant hand, leads to easier control than combining the two in one gesture command. Basic TV controls are more simple and do not necessarily require both hands for execution. Larger interaction space in front of the wall-sized display could also entice utilization of both hands.

Users are also more likely to come up with gestures which are depictions of the referent such as metaphoric, symbolic or iconic gestures, or utilize conventionalized, communicative gestures such as semaphorics [Wu and Wang, 2013; Aigner et al., 2012]. Whenever a task is too abstract to be expressed by a single gesture or a gesture phrase, it may be more appropriate to use widgets on the screen which are simply manipulated with pointing gestures [Vatavu, 2012]. Vatavu and Zaiti [2014] found that users emphasize either hand posture or hand movement in their elicited gestures but rarely these two properties are utilized simultaneously. Directly mapped gestures are easier to memorize and referents with opposite effects should have similar gestures [Wu and Wang, 2013; Vatavu and Zaiti, 2014]. For example, to increase volume a gesture that points up should be utilized and vice versa, a gesture pointing down should decrease volume. Findings of Nancel et al. [2011] also suggest that linear gestures lead to more accurate and faster performance. For example, zooming is controlled by moving an arm back and forth in front of a display rather than with more complex gesturing, with circular movement for instance.

One interesting finding yielded from elicitation studies [Vatavu, 2012; Vatavu and Zaiti, 2014] is that users often fall back on previously acquired interaction methods. This tendency can be seen as a preference for 2D interaction and less frequent exploitation of the depth dimension. Perhaps introduction of a third dimension is disorienting and complex. If depth information does not add value to interaction, one needs to consider not implementing it in the first place.

User-created gestures often mimic interaction techniques for touchscreen devices or desktop GUIs. Swipe, tap and pinch gestures are regularly suggested. An intriguing observation is that it appears that users tend to approach mid-air gesture interfaces by imagining an invisible 2D plane in front of them and interact on it as if it was a touchscreen. Another interesting approach is to draw letters in mid-air to invoke an event in a similar fashion to shortcut keys are used in traditional GUI applications. For example, a task ”Open Menu” is identified by drawing a letter M in the air. In addition, some users imagine a tangible object such as a turning button and act as if they were actually fiddling with it. Thus, it might be suitable to consider building upon the conventional interaction methods and refine older and familiar strategies in a way that fits the interaction in natural user interfaces. Of course, NUI technology should expand the repertoire of interaction strategies and enhance the performance in contexts where appropriate but reliance on more familiar methods would help adopt the new style of interaction.

Although it seems more suitable to let users propose gesture commands, elicitation approach has drawbacks. Studies have shown that there is a relatively low consensus among participants regarding elicited gestures and their expected effects. Studies have yielded average agreement scores between 20% and 40% [Wu and Wang, 2013; Vatavu, 2012; Vatavu and Zaiti, 2014; Pyryeskin et al., 2012]. A couple of reasons for this might exist. First, perhaps users become too creative and suggest gestures that are too complex, abstract or counterintuitive and ignore their natural first guess. Second, cultural and linguistic backgrounds shape gesture vocabularies even though Aigner et al. [2012] do not believe that major differences could be observed. Individual gestures may be different but they usually belong to the same gesture category despite the user's cultural background. For example, in Western cultures hand is extended with palm facing towards to indicate 'stop' sign. Japanese equivalent of this is to cross arms in front of the upper body. Both of these gestures, however, belong to static semaphoric gesture category.

In relation to the discussion about user-defined gestures, one has to consider the proposition of Zamborlin et al. [2014] to offer fully tailorable gesture sets. Users would then become the designers. This approach would allow more control and freedom for the user but at the same time usability features would be violated. Findings from the study of Vatavu and Zaiti [2014] do not support this approach. Interestingly, the recall rate for the participant's own gestures was only 72.8%, in 15.8% of all cases participants replayed a wrong gesture and in 11.4% of all cases participants could not remember the gesture they had just created. In addition, Pyryeskin et al. [2012] compared designer-created and user elicited gesture vocabularies. Results from their study suggest that designer-chosen gestures could lead to better performance and usability than gestures created by the users themselves.

2.5.3 Gorilla-arm effect

Another issue that deserves to be mentioned is the design of ergonomic gestures. When gesturing in mid-air users often report fatigue and a feeling of heaviness in the upper arm, a condition commonly known as the gorilla arm effect. Some design guidelines have been introduced in order to reduce arm fatigue. According to Hincapié-Ramos et al. [2014] the least amount of endurance is consumed when the arm is bent and interaction takes place midway between shoulder and waist line. Gestures which require arm movements above shoulder height are the worst. Muscular contraction is high and these gestures can be maintained only for a short period of time before energy is used up. Therefore, downward and horizontal movements should be preferred [Hincapié-Ramos et al., 2014; Wu and Wang, 2013]. Whenever possible, interface objects should be located closer to the bottom of the screen and commands should make use of horizontal movements such as swipe gestures.

According to Wu and Wang [2013] static, small-scale movements are perceived more comfortable than dynamic, complex or large-scale movements. Hincapié-Ramos et al. [2014] suggest that freehand interfaces should enable relative movements which means that the user is not forced to execute a gesture in a fixed position in the air.

Possibility to switch seamlessly between hands could be one way to reduce fatigue. Providing guidance to use dominant hand for certain tasks and secondary hand for simple and less frequent tasks could be one way to avoid gorilla-arm effect. In the experiment of Nancel et al. [2011] two-handed techniques were considered less tiring compared to one-handed techniques.

It should be kept in mind that freehand interaction is the most suitable for fast interactions. Whenever task requires continuous manipulation or maintaining a pose for a long time, alternative types of human-computer interaction should be considered. Perhaps the key to the design of ergonomic interface is the versality of available gestures. The number and frequency of bent and extended, static and dynamic gestures should be balanced in a way that is appropriate for the context interaction takes place in.

2.5.4 Learnability of gesture interaction

In the previous two sections the focus was on the design of gesture commands. In this section, subject is expanded to the design of mid-air interaction itself. The goal is to explain a few key concepts for creating a smooth novice-to-expert transition and the type of interaction that can easily be adopted. The concepts and strategies presented here are not new and the very same concepts have been utilized in other types of interaction design. Fundamentals are the same but they are manifested in different kind of actions.

New methods require learning and training. Maike et al. [2014] also address this issue in their heuristics. Two of these are obvious. *Learnability* heuristic guides the designer to balance learning time and frequency of use. If the usage is frequent, it is then acceptable to take longer to learn the technique. With a *learning* heuristic Maike and others also emphasize the importance of sociality and learning by copying actions of other users. There are a few techniques how to achieve good-quality training of users.

One of these techniques is scaffolding. According to Wigdor and Wixon [2011, p. 53] scaffolding is ”the creation of a design that promotes autonomous learning by employing actions that encourage users to develop their own cognitive, affective, and psychomotor skills”. For them, scaffolding is a concept that enfolds all the key elements in the design of an interaction that leads to successful transition from novice to expert.

One way to implement scaffolding is with a step-by-step strategy. The idea is to break the whole interaction into smaller and simpler actions. Using cues and hints embedded in the interface elements themselves the goal is to free users from memorization of technique and from the endless possibilities for an action. The user is led to the next action instead. Scaffolding supports learning by doing and exploration of the possibilities for actions. But the user cannot know what to do if the interface does not ”afford” these actions.

As an example from GUI context, buttons afford clicking through their shape and resemblance to the real-world objects. Since the object is not real but virtual, design has to rely on a user understanding that clicking the object is the correct and meaningful action to be performed. Norman [2004] referred to this concept as *perceived affordance.*

The concepts of scaffolding and perceived affordances are present in the self-revelation technique. In his work with marking menus for pen-based input, Kurtenbach [1993] introduced the concepts of self-revelation, guidance and rehearsal. *Self-revelation* means providing information to the user about available commands and how to invoke those commands. *Guidance* is to provide information while invoking a command and support the user in a completion of command. The goal of *rehearsal* is to teach through guidance how to physically invoke commands the way expert users would do. This way a smooth transition from novice to expert behavior can be achieved.

Walter et al. [2013] compared strategies for revealing an initial gesture command for their interactive public display game StrikeAPose. The gesture to execute was a teapot gesture. Three strategies were implemented. In spatial condition the screen was split into game area and ribbon below explaining the gesture with text, icons and a video. Temporal strategy was to interrupt the game for a short amount of time to show how to perform the teapot gesture in a video in the center of the screen. Integrative approach used three kinds of cues embedded in the game itself. A virtual user performing an example of gesture execution, mirror image of the user temporarily dispossessed of the user's control and showing the gesture or placing a button at the hip of a user's contour image to afford users to touch their hip.

Field study results show that spatial display of a gesture was the most effective strategy since 56% of the interacting users performed the gesture. The rate was significantly better than with the integration strategy with which 39% of the users executed the gesture. The temporal strategy was close with 47%. With the temporal and integrative strategies people also gave up more quickly and left whereas correct gesture execution took longer with the spatial approach but people were not so easily disengaged.

Sodhi et al. [2012] approach self-revelation from a different angle. *LightGuide* is a system that projects guidance hints directly on the user's hands. Arrows, hue coloring and predictive 3D pathlets are used to provide cues about the direction of hand movements. Compared to video instructions, users performed gestures nearly 85% more accurately with LightGuide.

Besides self-revelation, scaffolding and affordances, the use of interaction metaphors is recommended. In the creation of an interaction metaphor one needs to be careful, though. If it fails, interaction will become confusing and the requirement of natural interaction will not be achieved.

As an example of metaphorical design Song et al. [2012] have based their whole system design on a skewer metaphor. Their Kinect-based implementation is intented for 3D virtual object manipulation tasks. Using a skewer metaphor translating, rotating or aligning objects bimanually is intuitive and easily understood.

What is in common for all the techniques presented here is that they provide support for active exploration and aim for the development of skilled practice. Important is also that training of users should not be divided into novice and expert phases but learning ought to be achieved through active performance.

Wigdor and Wixon [2011] point out that metaphors, methods and elements can be borrowed from GUI interaction but at the same time a warning is sent to NUI designers. If the old methods are copied as they have become known in current types of interaction then the end result may just be another GUI implementation. Only distinction is that the input device is worse.

2.6 How freehand gesture interfaces compare to other interface types?

In this section I shortly present results from a few studies that have compared freehand gesture control to other types of input such as remote device-based control and touch input. The focus is on quantitative task performance and user satisfaction. Results from the studies presented here have not found convincing evidence for the efficiency of gesture-based control. User acceptance instead has produced slightly more variety.

Cox et al. [2012] compared input techniques for interactive TV applications. Microsoft Kinect, Wiimote and two methods using Android tablet were used. Participants conducted navigational tasks which included text entry and drag-and-drop tasks. Results show that users with Kinect had the lowest number of successful target hits in a drag-and-drop task indicating lower speed and accuracy. Kinect also had the highest error rate. In a text entry task Kinect was the slowest and less accurate input modality.

Prior experience had an effect, but still, expert users with Kinect did not achieve performance level nearly as good as with other devices. Despite seemingly poor performance of Kinect several participants enjoyed freehand interaction and liked the concept. Even 13% of the participants thought that this type of interaction could be useful. Still, Kinect was the least liked of all the techniques.

Bobeth et al. [2014] also investigated different input modalities for iTV applications. Three input techniques (remote control, tablet control, freehand gesture control) were compared. Experimental tasks were related to the usage of two iTV applications. Performance with remote control and freehand gesturing enabled by Kinect sensor were slower than tablet interaction. Freehand interaction was also rated lowest on all of the user experience dimensions (overall usability, effectiveness, satisfaction, efficiency) and it was the least preferred choice of input modality.

Heidrich et al. [2011] do not provide support for gesture control either. In their study of analysing different input technologies for interacting with smart wall they compared direct touch control, remote trackpad control and remote gesture control. Performance was assessed subjectively and quantitatively. Tasks involved using a healthcare application and a Fitts' law task. Freehand and trackpad were slower than direct touch control. Remote gesture control also caused more strain on the arm and shoulder. Not only did it physically burden the user, gesture input was also rated highest in cognitive effort.

Despite the fact that experimental results do not confirm the benefits of freehand interaction, there are still situations where this kind of interaction will be useful. In Section 2.4, a few examples from application domains were introduced. For instance, sterility can only be achieved with freehand interfaces. Also, controlling wall-sized displays directly from a close distance seems inappropriate. Gestures enable manipulation from a distance for which the large displays are intended for. Lastly, novel methods are often tested in traditional use cases. With methods primarily suitable for freehand interaction, gesture interfaces may turn out to be much more efficient but much depends on the imagination of developers who invent new techniques. Suitable gesture commands can be sought in elication studies or as an option, developers could use software especially intended for interaction design. Two of these are presented in the next section.

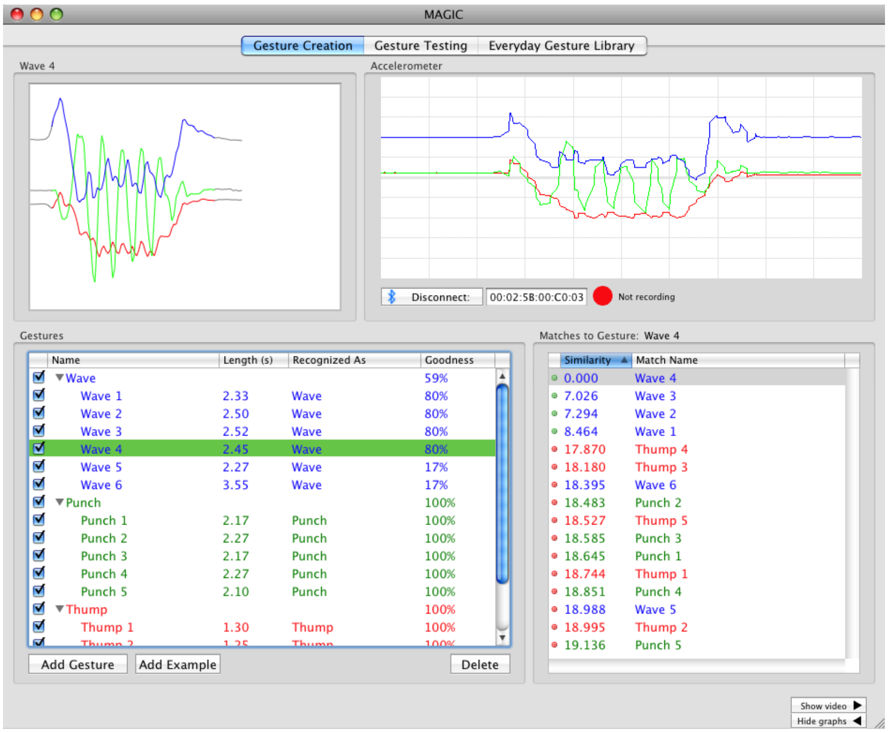
2.7 Tools for interaction design

Elicitation studies are not the only way of constructing a gesture vocabulary and designing gesture commands. Software tools for interaction design have been developed. Here two examples are presented.

Ashbrook and Starner [2010] have developed a software for gesture interaction designers. MAGIC stands for *Multiple Action Gesture Interface Creation* and it has been developed to solve two design-related issues. One is to offer a tool for interaction designers who are not experts in pattern recognition. The other is to provide a testing tool that would aid in searching for commands that would be different from the user's everyday gestures.

The tool is intended for finding meaningful gesture-functionality mappings, ensuring they work properly and testing that gestures work in conjunction with the user's natural movements. It is not used for gathering design requirements or final user testing.

MAGIC supports flexible three-stage workflow. These stages are gesture creation, gesture testing and false positive testing. In the first stage, the designer creates gesture classes and gesture examples. A class represents one kind of movement and several examples are created for each class. Examples are video recorded and data about recognition performance is shown in the interface (Fig. 4). In the second stage, designer tests the gesture classes and examples by performing gestures as they are intended and making motions that could be falsely recognized as a gesture command. In the last stage, created gestures are compared to the pre-recorded gestures in Everyday Gesture Library to find potential actions that could confuse gesture recognition.



Gestures

	Name	Length (s)	Recognized As	Goodness
☑	▼Wave			59%
☑	Wave 1	2.33	Wave	80%
☑	Wave 2	2.50	Wave	80%
☑	Wave 3	2.52	Wave	80%
☑	Wave 4	2.45	Wave	80%
☑	Wave 5	2.27	Wave	17%
☑	Wave 6	3.55	Wave	17%
☑	▼Punch			100%
☑	Punch 1	2.17	Punch	100%
☑	Punch 2	2.27	Punch	100%
☑	Punch 3	2.17	Punch	100%
☑	Punch 4	2.27	Punch	100%
☑	Punch 5	2.10	Punch	100%
☑	▼Thump			100%
☑	Thump 1	1.30	Thump	100%
☑	Thump 2	1.25	Thump	100%

Matches to Gesture: Wave 4

Similarity ▲	Match Name
0.000	Wave 4
7.026	Wave 3
7.294	Wave 2
8.464	Wave 1
17.870	Thump 4
18.180	Thump 3
18.395	Wave 6
18.483	Punch 2
18.527	Thump 5
18.585	Punch 3
18.645	Punch 1
18.744	Thump 1
18.851	Punch 4
18.988	Wave 5
18.995	Thump 2
19.136	Punch 5

Fig. 4. MAGIC. 'Gesture creation' tab open.

Zamborlin et al. [2014] have developed Gesture Interaction DEsigner (GIDE) (Fig. 5) which is a gesture recognition application meant to work across different application domains and media. While MAGIC is a software intended for expert developers, GIDE is a gesture design tool for actual users. GIDE supports four properties of gesture interfaces discussed earlier in Section 2.5.1 regarding design heuristics. These properties were continuous control, tailorable for specific context, meaningful feedback and allow expert and non-expert use.

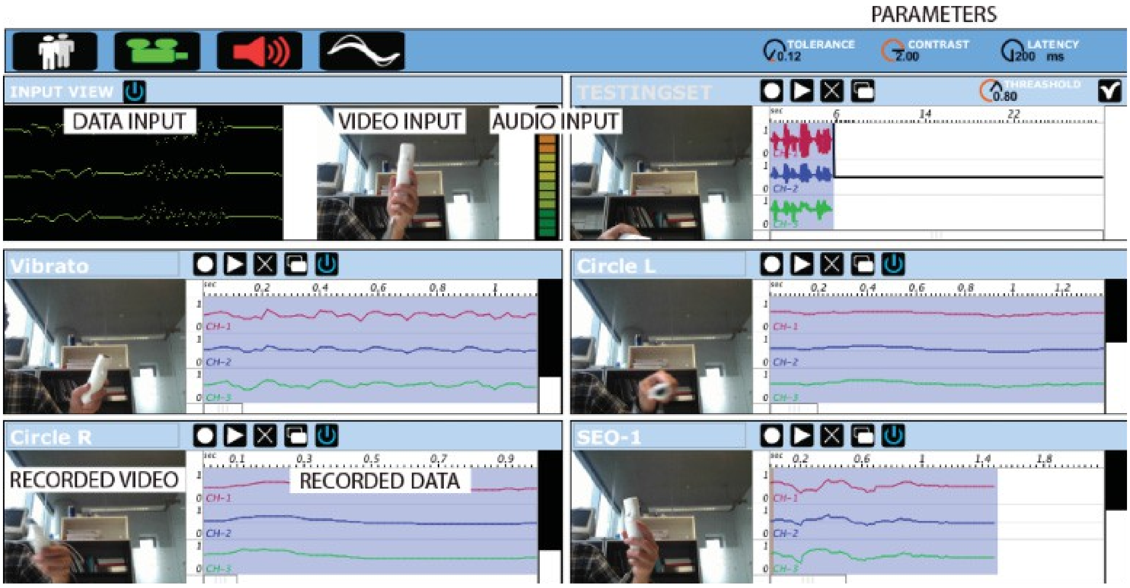


Fig. 5. GIDE application.

Like MAGIC, GIDE follows a three-phase iterative workflow. Phase one is recording a gesture. Recorded gestures can be edited and this way users can easily build their own modified gesture vocabulary. Phase two is named ”Follow” mode and real-time feedback. In the follow mode users perform a gesture and the application gives a moment-by-moment probability estimation of the gesture being performed and the phase of the gesture. For instance, recorded audio can be attached to the recognition of a gesture. Batch testing is also supported within phase two. Third phase is tuning the parameters of the machine learning algorithm. The user determines how much the performance is allowed to be different from pre-recorded gestures by tuning the tolerance parameter. Latency of the system can also be modified and it basically tunes a balance between the reaction speed to input and the reliability of gesture recognition. Contrast changes probability values of a gesture in the vocabulary. Higher the normalized contrast parameter, higher the difference between gestures.

2.8 Summary

In this chapter freehand gesture interaction was discussed. The chapter started with a definition of gesture. A shared view is that gestures are movements that convey meaning. One issue in which definitions differ from each other is whether this meaning can be expressed through non-verbal gestures alone or always in concert with speech.

Several gesture classifications for human-human and human-computer interaction were presented. Independent of the context in which these classifications were constructed, most of the categorizations share underlying similarities. Categorizations proposed by Karam and Schraefel, Efron, McNeill and Kendon are mostly referred in this work. In the experiment that is described later in Chapter 4, pointing and manipulative gestures are investigated in a simple data entry task.

The naturalness of gesture interaction was also discussed. The end result of contemplation was that labeling gesture interfaces as natural might be misleading. Interfaces are never inherently natural but they can become natural through meaningful actions they enable and the feeling of naturalness can be strengthened through learning and exploration.

Three examples of application domains in which gesture interaction has been shown to be beneficial were introduced. The most obvious advantage of freehand gesturing is that it enables sterile interaction which is a crucial requirement in operating rooms for instance. Public and wall-sized displays as well as iTVs have been studied extensively in the area of gesture interaction. Two examples from these domains were presented.

Four sections were dedicated to the design issues. First, heuristics for the design of gesture interfaces were presented. Furthermore, heuristics for traditional GUI design and gesture interfaces were also compared. Already established rules of thumb for GUI design should not be in any way neglected but new issues such as emphasis on tracking accuracy should be taken into account in the development of modern gesture interfaces.

A few things should be considered in creating a gesture command. One-handed commands are preferred unless two-handed gesturing is necessary to execute the function. Whenever a task is not too abstract, gesture command should depict an icon of the referent or emblems should be used. Moreover, users remember commands which are directly mapped to functions they represent. Introduction of a depth dimension should be pondered since it could possibly confuse users. Ergonomic factors should also be considered to reduce fatigue. The most optimal gesture command is executed between shoulder and waist line with a bent arm. In addition to the design of a single gesture command, interaction in a wider perspective was discussed. Scaffolding, perceived affordances, self-revelation techniques and the use of metaphors were offered as solutions to enhance the learnability of freehand gesture interaction.

After design issues, gesture interfaces were compared to remote control and touch-based interaction in terms of task performance and user satisfaction. Freehand gesture input method may not achieve as efficient performance as touch-based interaction or remote control in certain tasks that are not specifically intended for gesture interaction. However, there are advantages that cannot be achieved with any other type of interface. For instance, the requirement of sterility can only be fulfilled with contactless interaction. Furthermore, perhaps the utilization of haptic feedback could improve performance with freehand techniques. The next chapter scrutinizes the benefits of haptic feedback from different perspectives.

At the end of the chapter software tools for interaction design were presented. Software tools like MAGIC support not only the design of an alternative command but also fine tuning of the gesture.

3 Haptics

Freehand gesture interaction has one major disadvantage. It lacks passive feedback and users can only rely on proprioceptive feedback. Nancel et al. [2011] use the term *degree of guidance* to describe the trade-off between passive feedback that is received through actually touching the device and the available degrees of freedom of the device. One-dimensional devices provide the greatest amount of passive feedback but can only allow restricted movement. For example, a mouse wheel allows movement only on one axis. Touch-sensitive surfaces are two-dimensional but possess limited guidance by haptic feedback. Mid-air gesture interaction provide multiple degrees of freedom but passive feedback is absent. Haptic feedback is important in providing a sense of direct interaction and control. Due to the fact that touch is so inherent to us, haptic feedback is also essential in making the interface feel natural.

In this chapter, haptic feedback in gesture interaction is studied. First, I shortly define haptics and explain which mechanoreceptors are responsible for different properties of touch sensation. After that I offer evidence for and against a benefit of haptics in quantitative task performance, multimodal and non-visual interaction, haptic guidance and user satisfaction. At the end of the chapter, I go through techniques and technological devices developed for the implementation of haptic feedback. In addition to contactless feedback technologies I present touch-based techniques.

3.1 *What is haptics?*

Haptics is divided into two main categories: kinesthetics and tactile feedback [Rovan and Hayward, 2000]. Haptics can also be separated into two subcategories by the nature of the haptic properties of an object. Whenever feedback is received through tangible interaction and produced by the physical properties of an object, haptic stimulation type is referred to as *passive haptics*. When feedback is generated by the device, it is referred to as *active haptics.*

Here I rely on definitions provided by Rovan and Hayward [2000] and Subramanian et al. [2005]. Kinesthetics focuses on limb movement and orientation of body parts. Sensory information is received via proprioceptors such as muscle spindles and Golgi tendon organ. Force feedback is also a tightly related term used to refer to information interpreted by muscular, skeletal and proprioceptive senses.

Tactile and vibrotactile feedback are often used interchangeably but the skin's sense of touch can interpret a variety of sensory information (e.g., texture, pressure, curvature and thermal properties). Tactile, or vibrotactile, feedback refers to sensory information received via cutaneous inputs such as mechanoreceptors that are specialized to certain stimuli types.

In this work, emphasis is on vibrotactile feedback. The subject is limited due to the fact that Leap Motion application utilizes this kind of feedback and its effects are studied in the experiment carried out for this thesis work. The application and the experiment are presented later in Chapter 4. In the next section, functions and properties of four mechanoreceptors are explained in more detail.

3.2 *Mechanoreceptors*

Perceptual process of touch is explained by four-channel model of mechanoreception [Bolanowski et al., 1988]. The model consists of four psychophysical (information) channels 1) P *(Pacinian)*, 2) NP I *(non-Pacinian)*, 3) NP II and 4) NP III and their neurophysiological substrates. Functions and properties of the information channels and the skin receptors are summarized in Table 4.

Channel	P	NP I	NP II	NP III
Afferent fiber type	PC	RA	SA II	SA I
Receptor	Pacinian corpuscle	Meissner corpuscle	Ruffini ending	Merkel cell
Rate of adaption	Rapidly adapting	Rapidly adapting	Slowly adapting	Slowly adapting
Receptive field	May include an entire hand	3-5 mm in diameter	1-2.5 cm in diameter	2-3 mm in diameter
Stimulus frequency	40-300 Hz	1.5-50 Hz	15-400 Hz	0.4-1.5 Hz
Function	Perception of high frequency stimulation	Low frequency vibration, skin slip	Skin strech, object motion and direction	Form and texture perception; points, edges, curvature; static pressure/indentation
Location	Deep, subcutaneous	Shallow, dermis	Deep, subcutaneous	Shallow, dermis

Table 4. Properties of information channels and receptors.

Four cutaneous mechanoreceptive afferent neuron types innervate the glabrous (non-hairy) skin of human hand. They can be categorised by their rate of adaption as slowly or rapidly adapting. Pacinian corpuscle (PC) fibers that end in Pacinian corpuscle receptors provide input for the P channel and rapidly adapting (RA) fibers that terminate in Meissner corpuscles are the physiological correlates of NP I channel. PC fibers and RA fibers both belong to rapidly adapting category. Slowly adapting type I (SA I) fibers which innervate Merkel cell receptors *(also Merkel neurite complex or Merkel disk)* are the neural inputs for NP III channel. NP II channel is a psychophysical correlate for slowly adapting type II (SA II) fibers that end in Ruffini endings *(also*

Ruffini corpuscle or SA II end organ). [Bolanowski et al., 1988; Johnson, 2001; Gescheider et al., 2002]

In addition to rate of adaption, afferent fibers can be categorised by the size of their receptive field. PC and SA II fibers have large receptive fields but low spatial resolution. The receptive field of Pacinian afferents may include an entire hand and SA II type afferents have receptive fields of 1-2.5 cm in diameter. SA I and RA afferents have smaller receptive fields but their spatial resolution is higher. SA I afferents have a receptive field of 2-3 mm in diameter but they are capable of resolving spatial detail of 0.5 mm. The receptive field of RA afferents is 3-5 mm in diameter but they respond to stimuli over the entire area and thus resolve spatial detail poorly. [Johnson, 2001]

Meissner corpuscles and Merkel disks are located in the upper layers of the dermis, close to the basal layer of the epidermis. Pacinian corpuscles and Ruffini endings are located deeper in the subcutaneous tissue beneath the dermis. [Wu et al., 2006; Johnson, 2001]

Operating range of the four information channels for the perception of vibration is between 0.4 Hz and 500 Hz [Bolanowski et al., 1988]. Sensitivities may overlap and perceptual qualities of touch may be determined by the combined inputs from four channels [Bolanowski et al., 1988]. Afferent fibers and receptors contribute differently to perceptual process of touch and they are specialized to operate at certain frequency ranges and detect certain stimuli.

Pacinian corpuscles are sensitive to high-frequency vibration and responsible for the perception of distant vibrations transmitted through an object held in the hand [Johnson, 2001]. Operating range of PC corpuscles is between 40 and 300 Hz, peak values reached around 125-250 Hz above which frequency sensitivity decreases substantially [Wu et al., 2006].

Operating range of vibration frequencies for NP I channel falls between 1.5 and 50 Hz [Gescheider et al., 2002]. Meissner corpuscles are especially responsible for the detection of low frequency vibration [Johnson, 2001]. Detection threshold of NP I channel is optimally tuned at 30-50 Hz [Gescheider et al., 2002]. Meissner corpuscles are insensitive to static force but capable of detecting gaps in a grating but only until they are wider than the receptive field of Meissner corpuscle which is 3-5 mm in diameter [Johnson, 2001]. Due to this low spatial acuity, RA afferent fibers are sensitive to detecting slips between the skin and an object held in the hand [Johnson, 2001].

Vibration-frequency range for NP II is from 15 to 400 Hz [Bolanowski et al., 1988]. Although frequency range is largely similar with P channel, NP II channel operates at much lower sensitivity [Bolanowski et al., 1988]. Ruffini endings are responsible for the detection of skin stretch and they are also involved in the perception of object motion, direction and orientation [Johnson, 2001].

NP III channel detects very low frequency stimuli from 0.4 to 1.5 Hz [Gescheider et al., 2002]. Merkel disks are responsible for texture and form perception and they are especially sensitive to points, edges and curvature much more than to indentation as such [Johnson, 2001].

3.3 Evidence for and against haptic feedback

In this section I examine the usefulness of haptic feedback in general. Evidence from the experiments studying mid-air gesture interaction as well as mobile and GUI interaction is provided. Findings from the studies examining quantitative task performance, multimodal interaction and non-visual interaction are presented. Furthermore, results regarding haptic guidance and user satisfaction are discussed.

3.3.1 Quantitative task performance

Overall, the majority of studies have confirmed that haptic feedback can enhance performance significantly. Passive haptic feedback has been confirmed to enhance performance since higher degree of guidance results in more accurate and more efficient performance [Nancel et al., 2011]. Better performance results have also been verified in the studies of active haptic feedback. Nevertheless, mixed results and contrary evidence can also be found in the literature.

The advantage of haptic feedback has already been confirmed in traditional GUI interaction. In their study, Dennerlein et al. [2000] investigated how a force-feedback mouse could improve movement time performance compared to a conventional mouse. The experiment featured a steering task and a combined steering-targeting task. The force-feedback mouse produced force that pulled the cursor to the center of the target tunnel. When force-feedback was enabled, movement time was on average 52% faster for a drag task and 25% faster for a drag-and-drop task.

The benefit of haptic feedback has also been proven in mobile interaction. Brewster et al. [2007] investigated the use of vibrotactile feedback for touchscreen keyboards on PDAs. A vibrotactile actuator was placed to the back of the PDA and it generated two different stimuli to indicate a successful button press or an error. Text entry task was performed in a laboratory setting and on an underground train. The results show that tactile feedback significantly improved task performance. In a laboratory setting more text was entered, fewer errors were made and more errors were corrected while vibrotactile feedback was enabled. In a mobile setting tactile feedback was less beneficial as the only significant difference was found in the number of errors corrected.

In another experiment finger-based text entry for mobile devices with touchscreens was studied by Hoggan et al. [2008]. A physical keyboard, a standard touchscreen and a touchscreen with tactile feedback added were compared. Tactile feedback constituted of a set of tactons which represented different keyboard events and keys. Text entry tasks were performed in the laboratory and mobile settings (on the subway). The results show that participants entered text more accurately, with lesser number of keystrokes per character and faster with a physical keyboard. Tactile condition was close to the performance of a physical keyboard and significantly better than standard condition. Overall workload was significantly higher when participants used standard touchscreen than either physical or touchscreen with tactile feedback. In addition, customized version of the tactile feedback was also tested. Two vibrotactile actuators were placed on the back of a PDA device to provide more specialized feedback and it was found that performance could be improved even further using more accurate and specified feedback.

Positive findings have also been found in the studies of mid-air gesture interaction. Adams et al. [2010] investigated the effects of vibrotactile feedback in mid-air gesture interaction. Participants performed basic text entry task on a virtual keyboard with and without the tactile feedback. A vibrotactile actuator was located on the index finger inside a glove and feedback was generated to indicate a positive confirmation of the keystroke event. It was found that participants entered text significantly faster with tactile feedback than without it. No differences between the conditions were found in error rate and keystrokes per character measurement.

Krol et al. [2009] compared visual, aural and haptic feedback types in a simple remote pointing task. The experiment involved a two-dimensional Fitts' law task with circular targets. In terms of movement time and time on target haptic feedback significantly improved performance compared to visual feedback alone. No significant differences were found between haptic and aural conditions. However, error rate per participant was the worst in haptic condition, aural condition being slightly better and visual condition having clearly the least number of errors.

In contrast to the previous results, Foehrenbach et al. [2009] did not confirm a benefit of tactile feedback on user performance. They studied hand gesture input in front of a wall-sized display with and without tactile feedback. The experiment featured one-directional Fitts' tapping tasks. Continuous vibration was generated by shape-memory alloy wires attached around three fingertips inside the data glove markers. The results show that non-tactile feedback performed slightly better in terms of throughput and movement time although no significant differences were found. Also horizontal and vertical target alignments were compared and significantly higher error rate was found for the horizontal alignment when tactile feedback was used.

Studies presented here only considered comparisons between visual and haptic conditions. When an additional auditory modality has been included, studies have generated intriguing results about the advantages of haptic feedback.

3.3.2 Haptic feedback in multimodal interaction

Experiments studying multimodal interaction have revealed interesting facts about the characteristics of haptic modality. It has been substantiated that visual feedback alone is inadequate but an additional auditory or haptic modality benefits interaction differently.

In a meta-analysis Burke et al. [2006] compared visual-auditory and visual-tactile feedback to visual feedback alone and examined the effects on user performance. Study revealed that adding an additional modality to visual feedback enhances overall performance. However, the advantages of additional modalities are different. Whereas visual-auditory feedback is the most efficient when a single task is performed and under normal workload, haptic feedback is more effective with multiple tasks and when the workload is considered high. Visual and auditory feedback types seem to increase experienced workload. Both conditions produced favorable performance in target acquisition tasks but tactile feedback was beneficial for alert, warning and interruption tasks for which auditory feedback was not effective. Neither one of the multimodal conditions were effective in reducing error rates.

Some studies have not found confirmation for a benefit of haptic feedback. The study of Jacko et al. [2003] yielded results which do not fully support the argument that haptic feedback as a sole additional modality would improve user performance. Uni-, bi- and trimodal (visual, auditory, haptic) conditions were examined in a drag-and-drop task. A force feedback mouse provided mechanical vibration and a sound that resembled a suction cup was used as an auditory icon. Participants were older people (54 years and above) and either visually impaired with varying visual acuities or normal-sighted. As expected, visual feedback alone performed worst compared to multimodal conditions. Within all the groups additional auditory component appeared to enhance performance. A benefit of haptic feedback as a single additional modality was not confirmed but advantages were perceived when auditory component was involved.

Foehrenbach et al. [2009] suggest that tactile feedback can interfere with other senses in a negative way. They argue that visual and haptic information is delivered through different information channels and thus resulting in a lag in velocity of processing and reaction time. Asynchronous information processing can arouse irritation in a user and decrease the pleasantness of using the interface.

Despite distinctive advantages of different modalities, environmental factors often determine suitability of a feedback type. Hoggan et al. [2009] have shown that significant decreases in performance for audio feedback appear at noise level of 94dB and above while performance for tactile feedback starts to decrease at vibration level of 9.18 g/s.

There are situations where receiving visual and auditory feedback is not meaningful. The user's impairment in sight or hearing or environmental factors may limit the use of different modalities. In the next section, haptics in non-visual interaction is investigated in more detail.

3.3.3 Non-visual interaction

Mainly positive results have been found in the study of haptic feedback in non-visual interaction. Findings presented here are gathered from the studies involving direct interaction but they can also be applied to freehand interaction.

Charoenchaimonkon et al. [2010] compared audio and vibrotactile feedback methods in their study of non-visual pen-based input. Participants conducted a number of Fitts' law target selection tasks with varying levels of difficulty. Expectedly performance was slower and more error-prone in both audio-only and tactile-only conditions compared to visual condition. Overall, participants performed better with tactile feedback compared to audio feedback and the advantage increased as the size of the targets and a distance between them increased.

Charoenchaimonkon and others utilized feedback conventionally to indicate positive confirmation. Tactile feedback has turned out to be appropriate in providing negative feedback, that is, alerting users on faulty or ineffective actions.

Martin and Parikh [2011] studied the effectiveness of negative feedback in a teleoperated robot control task. Negative feedback informed users whenever an inactive part of the keyboard was pressed. The task was to navigate a robot through a maze. To steer the robot a numeric pad on a conventional keyboard and two soft keyboards on a mobile phone with or without tactile feedback were used. Results show that conventional keyboard outperformed both soft keyboards in terms of task completion time and number of times robot hit the maze wall. No difference was found in a number of keypresses between input devices. Comparing the soft keyboards data indicates a slightly better performance, although not significant, when negative feedback is enabled.

Tactile feedback has also been proven to be suitable for providing more complex information than just simple alerts or confirmations. According to Brown et al. [2005, p. 167] tactons are ”structured, abstract, tactile messages which can be used to communicate information non-visually” and they can be compared to earcons or visual icons. Tactons are structured by changing frequency, amplitude, duration and waveform of stimulation. Perceptually more distinctive tactons can be implemented by mixing different parameters. Additionally, rhytmic patterns are easily distinguishable. Brown and others have also studied the recognition of tactons. In their study, overall recognition rate of 71% for different tactons was achieved. For rhytmic patterns the recognition rate was as high as 93% and for roughness the recognition rate of 80% was achieved. These results suggest that tactons can be beneficial for communicating information in user interfaces.

The effectiveness of slightly different kind of tactile icons were studied by Pasquero and Hayward [2011]. A device called THMB (Tactile Handheld Miniature Bimodal device, pronounced *thumb*) which combines graphical and tactile feedback produced with piezoelectric actuators was used (described in more detail later in Section 3.4). The task was to scroll through a list made of numbers 1-100 and select a target item. The list was not visible and if participants needed to view the list they had to press and hold down a key. Two tactile icons were implemented. The other one was triggered for each item in the list, the other for every ten items in the list. Results show that the number of viewings was reduced by 28% when tactile feedback was enabled. Data also suggests that the addition of tactile feedback results in a less error-prone performance although this observation was not statistically significant. With tactile feedback, the time between two keystrokes and a number of overshoots were increased.

Negative feedback can be beneficial in freehand gesture interaction. For example, users can be informed of a failure in recognition or incorrect execution of a gesture command. Haptic feedback can also be provided if the system recognizes movement but cannot interpret it correctly. Another situation where vibrotactile feedback could be utilized is when the user's hand is not in the field of view of the device or hand is not detected accurately enough. As the user moves his arm, the distance from an optimal recognition location could be indicated with a rhytmic pattern of varying tempo. Further the hand is away from the recognition area, faster the tempo of the generated feedback.

3.3.4 Haptic guidance

The idea of haptic guidance is to direct a user towards the target or guide motion along the predefined trajectory. Here I present two experiments with contradictory results for a benefit of haptic guidance.

Lehtinen et al. [2012] have studied dynamic tactile cueing coupled with visual feedback in mid-air gesture interaction. In their experiment participants conducted visual search tasks in front of a large display. Raycasting technique was utilized for visual feedback and rich directional vibrotactile feedback guided the user by ”pulling” the hand towards the right target. The advantage of vibrotactile feedback was found but not consistently through conditions. Performance was increased especially within conditions where a number of visual items on the screen was large.

Results from the experiment conducted by Weber et al. [2011] do not fully support the advantage of tactile feedback in guidance. Weber and others compared two forms of directional tactile feedback and verbal instructions in non-visual mid-air interaction in which participants translated and rotated virtual objects. VibroTac on the participant's right wrist provided vibrotactile feedback with either four or six directional cues. Verbal feedback consisted of commands ”Up”, ”Down”, ”Right” and ”Left”. Results show that verbal cues produced faster task completion times. Workload for verbal instructions was also rated significantly lower and it was also rated more appropriate for guidance. The instruction method did not have an impact on the performance accuracy.

Although results indicate preference towards verbal instructions, haptic feedback has advantages that are difficult to achieve with an auditory component. Practically verbal guidance is limited to discrete directional commands whereas tactile feedback can, besides directional cueing, fluently offer continuous information about distance.

Earlier in the discussion of learnability of gesture interaction, Kurtenbach's concepts of self-revelation, guidance and rehearsal were introduced. At least in guidance and rehearsal of freehand gesture commands tactile feedback could be beneficial. Embedded into real-time gesture recognition, tactile feedback could indicate incorrect hand poses, trajectories or tracking errors proactively in the way as it was described at the end of the previous section.

3.3.5 User satisfaction

Subjective evaluations considering user satisfaction towards haptic feedback have produced results across a wide range of opinions. In a study by Brewster et al. [2007] subjects favoured the vibrotactile condition as it was viewed as being less frustrating and annoying. Vibrotactile feedback also reduced the overall workload. The addition of vibrotactile feedback can also create more natural feel to virtual objects and enhance the ease of use of an interface [Adams et al., 2010].

Experiment conducted by Foehrenbach et al. [2009] revealed an even preference towards tactile and non-tactile feedback. Some participants mentioned that they felt to be set under pressure due to the continuous tactile feedback. In addition, they speculate that some participants did not utilize tactile information but relied on visual feedback and just tolerated vibratory feedback. Perhaps continuous tactile feedback was not clearly bound to any specific event and therefore it did not capture the user's attention and the information it provided was not considered useful.

In a study conducted by Krol et al. [2009] none of the eight participants preferred haptic feedback over visual or aural feedback. In addition, haptic feedback was perceived as being the slowest of the three modalities although it was in fact the fastest. Krol et al. [2009] argue that the low acceptance rate might be caused by the sensory overload since the solenoid technology in a pointing device made a sound when actuated and thus visual, haptic and aural modalities were unintentionally mixed.

Perhaps differences in technical implementations and the quality of feedback could cause these diverse findings. Research has shown that people might perceive some haptic feedback types more pleasant than others. Koskinen et al. [2008] compared tactile feedback generated with piezo actuators and a standard vibration motor. There was a slight preference towards feedback produced with piezo actuators although the difference was not statistically significant. In any case, tactile feedback was superior to non-tactile condition regardless of the technology. Pfeiffer et al. [2014] evaluated preferences of electrical muscle stimulation (EMS) and vibrotactile feedback in freehand interaction. Results reveal that participants liked EMS more than vibrotactile feedback. Techniques were further evaluated with regard to virtual object properties such as soft or hard material. Participants considered EMS to provide more realistic feedback when interacting with virtual objects of varying properties (soft, hard, cold or pointed material).

3.4 *Tactile technologies*

A variety of technologies have been developed for the implementation of tactile feedback. In this section I briefly introduce only a few non-contact and touch-based techniques. Technologies are explained through examples and advantages as well as disadvantages are considered.

3.4.1 Vibrating motors

There are two common types of vibration motors. These are eccentric rotating mass motors (ERM) and linear resonant actuators (LRA).

Eccentric rotating mass vibration motors are DC-motors with a non-symmetric, off-center mass attached to the shaft (Fig. 6). As the shaft rotates, asymmetric force of the off-center mass results in a centrifugal force which causes constant displacement of the motor. This displacement is then sensed as vibration.

There are at least two disadvantages related to ERMs. One is their slow speed and response time. It takes some time to start and stop the rotation. The other one is the unability to manipulate waveforms with changes in amplitude levels. Only speed and frequency can be varied. Therefore, the generated feedback tends to be one-sided. ERMs are, nevertheless, an inexpensive option to be used.

Fig. 6. Eccentric rotating mass vibration motors. On the left, Xbox 360 vibration motors and on the right, motors used in vibrating toys. [Image source: http://openxcplatform.com/projects/shift-knob.html]

Linear resonant actuators consist of a wave spring, a magnetic mass and a voice coil (Fig. 7). When electrical current is applied to the voice coil, it creates a magnetic field that causes the magnetic mass to move towards the spring which returns it back to the centre. When this movement is repeated, vibration is generated.

Unlike with ERMs, more sophisticated vibration feedback can be provided by modifying the amplitude of an input signal. The response time of LRAs is also much faster.

Fig. 7. C2 Tactor from Engineering Acoustics Inc. is a linear vibrotactile actuator. [Image source: Eve Hoggan, Stephen A. Brewster, and Jody Johnston, Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 1573-1582.]

3.4.2 Solenoids

Lee et al. [2004] implemented Haptic pen (Fig. 8) which is a tactile stylus for touch screens. Haptic pen uses solenoid technology to create two types of feedback. The other one creates a sensation of stiffness when button is clicked. The other one produces buzzing feedback. A push-type solenoid at the eraser end of the pen moves up and down creating a kick that depends on the force of the pressure directed at the tip shaft. The microcontroller digitizes the pressure sensor and communicates with the PC which selects the appropriate feedback.

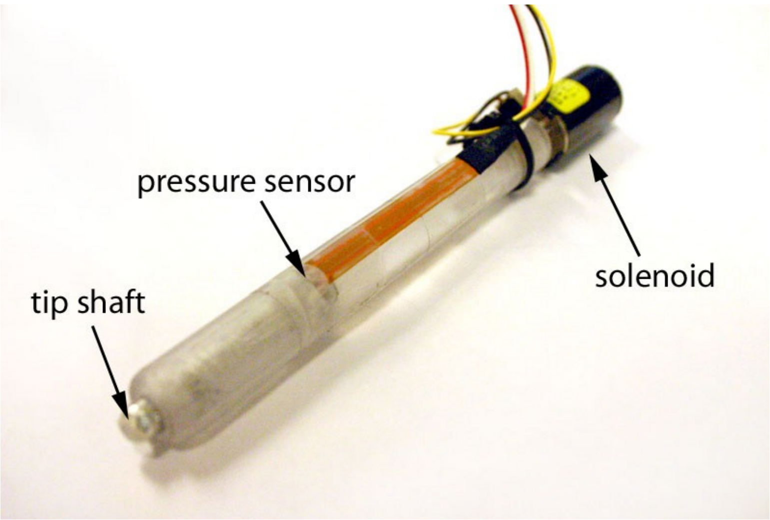


Fig. 8. Haptic pen.

A solenoid may be useful in situations where fast responses are required. Vibration motors may be too slow to start and stop so solenoids can offer feedback that is more exact. Solenoid technology is also inexpensive and the implementation is simple. Haptic pen also shows the technique's ability to create a relatively wide range of sensations regarding the straightforward implementation.

3.4.3 Electrovibration

TeslaTouch has been developed by Bau et al. [2010] and it utilizes the principle of electrovibration to provide tactile feedback in touch screens. Electrovibration is based on the electrostatic friction between a conductive surface and the skin. When alternating the voltage applied to the surface, electrically induced attractive force develops between a sliding finger and the underlying electrode. TeslaTouch uses specific panel that consists of an electrode sheet applied onto a glass plate which is covered with an insulator layer. Periodic electrical signal produces changes in friction and these changes cause skin deformations which in turn can be sensed as vibrations. The operating principle of TeslaTouch is depicted in Fig. 9.

Compared to electrocutaneous and electrostatic methods, electrovibration has a few advantages. No charge is passed through the skin as in electrocutaneous method so it is not an intrusive method and unlike electrostatic technique, electrovibration does not require an intermediate object to enable tactile sensing. Downside of electrovibration method is that feedback can only be felt when the finger is moving on the surface. However, electrovibration cannot be applied to mid-air gesture interaction although it is a scalable and efficient method to provide vibrotactile feedback and texture sensing for touch screens (Fig. 10).

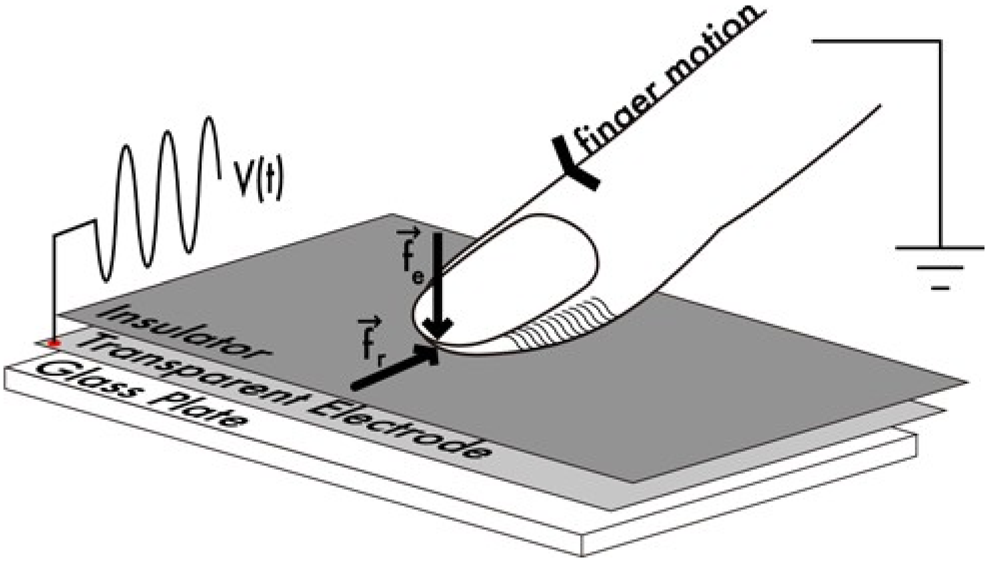


Fig. 9. TeslaTouch operating principle.

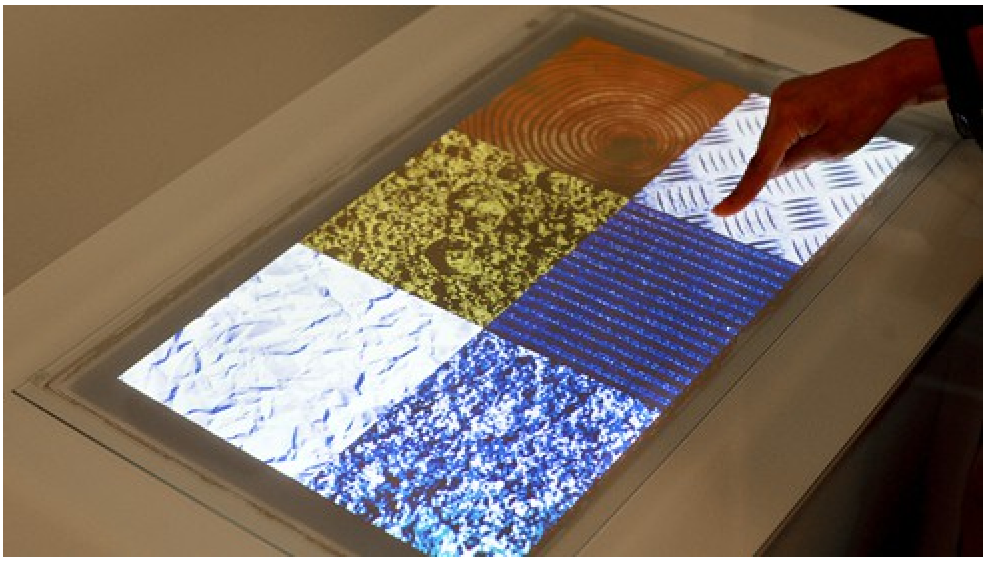


Fig. 10. TeslaTouch can produce a variety of textures.

3.4.4 Piezoelectric actuators

One way to produce tactile feedback is to use piezoelectric ceramic elements. Modulated voltage applied to the element is converted into a small mechanical bending motion. Utilizing this effect tactile feedback can be given directly onto the surface of the skin.

As was mentioned earlier in Section 3.3.3 which regarded non-visual interaction, Pasquero and Hayward [2011] have developed THMB which is a device that utilizes piezoelectric actuators (Fig. 11). Tactile stimulation is generated by using an array of eight 0.5 mm thick piezoelectric benders which cause deformations in the skin.

Piezoelectric actuators can be beneficial for systems in which fast reaction speed and low power consumption are required. When properly designed, using ceramic discs can also be a compact solution and still they can offer rich tactile feedback. In addition to plain buzzing feedback, utilizing stationary and independent deflections of ceramic discs sensations of shapes can be conveyed. Being a non-magnetic technology, piezoelectric actuators can be used in application domains such as industrial or medical applications.

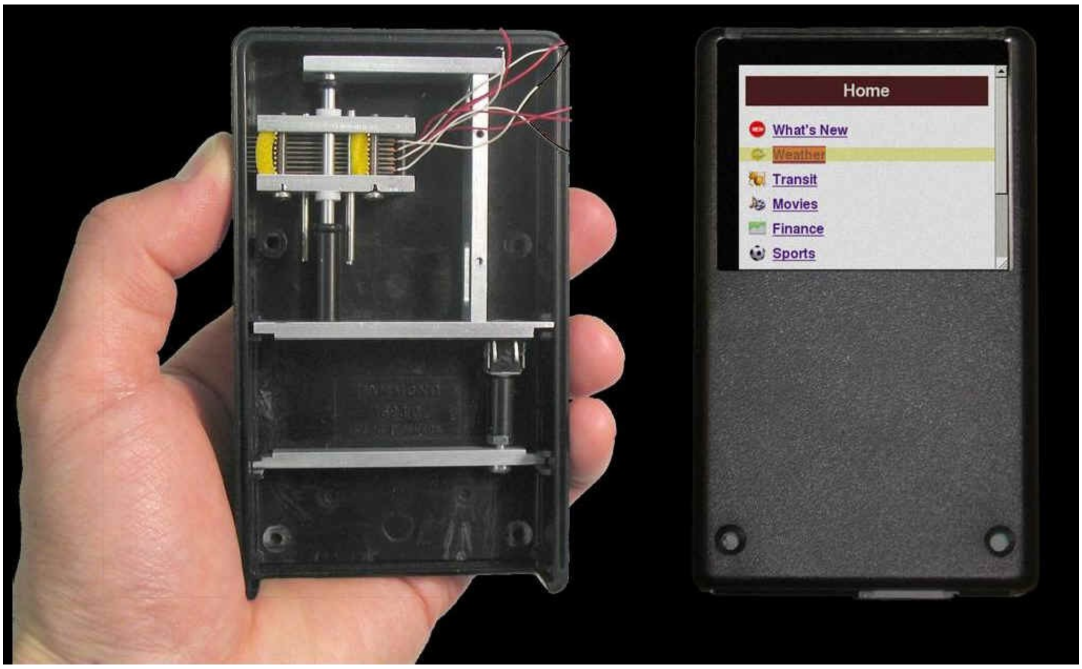


Fig. 11. THMB. [Image source: http://www.cim.mcgill.ca/~haptic/laterotactile/dev/thmb/]

3.4.5 Pneumatic systems

Here I present two pneumatic techniques. The first one utilizes air suction technique and the other generation of air vortices.

Designed for touch interaction, Hachisu and Fukumoto [2014] have developed VacuumTouch (Fig. 12). The system consists of an air vacuum pump, an air tank and an array of electric magnetic air valves connected to the holes on the surface. When the user's finger is located on a hole on the surface, an air valve is activated.

Air suction creates a sensation as if the finger would get stuck while moving it on the surface. Also, vibrotactile feedback can be produced although it may be weak.

VacuumTouch uses a 5 x 5 array of air valves that covers only a part of the surface, hence impairing dynamic feedback construction. However, using air suction is an interesting alternative for haptic feedback implementation. It is also possible to offer feedback above the surface with greater forces but only at near distances.

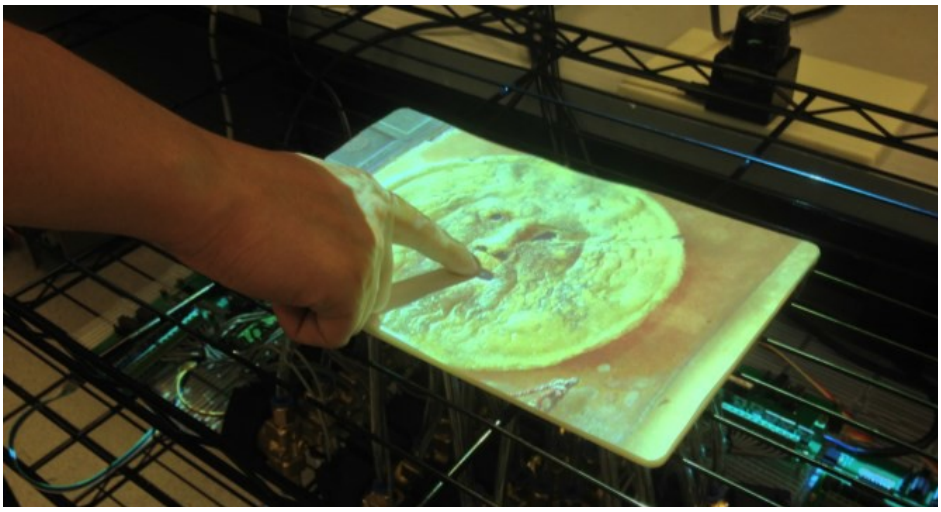


Fig. 12. VacuumTouch.

Sodhi et al. [2013] have implemented AIREAL (Fig. 13) system that makes use of vortices to provide haptic feedback in mid-air. AIREAL uses five subwoofers as actuators. These actuators contain a flexible diaphragm which quickly eject air out of a nozzle when displaced. The nozzle is directed towards the target within a 75 degree field of view. Manipulating the rate of displacement of the diaphragm different tactile sensations can be produced.

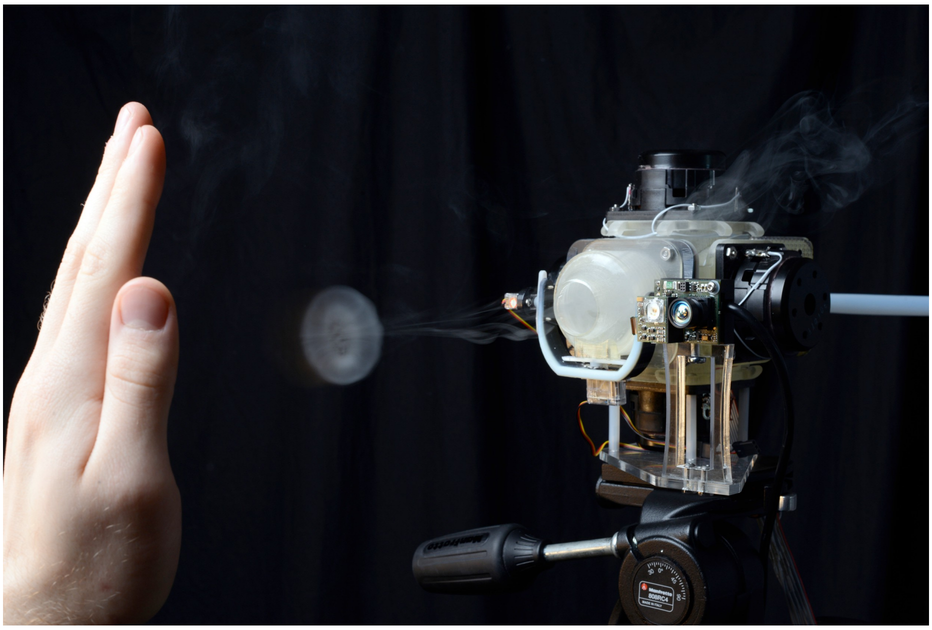


Fig. 13. AIREAL emits a ring of air targeted at the user's palm.

Ability to generate tactile sensations in free air is definitely an advantage. Moreover, feedback can be received from a distance. AIREAL is capable of providing effective feedback at one meter with a resolution of 8.5 cm. The accuracy of feedback at this distance is sufficient enough even though highly detailed feedback cannot be created.

Practical applications may require a number of devices placed at different spots which takes up a lot of space. Like the air suction technique, air vortex generation can also be noisy which diminishes the practicality of the technology. Nonetheless, the utilization of air vortices is a promising alternative for freehand gesture interaction. The other one being ultrasound which is presented next.

3.4.6 *Ultrasonic transducers*

Carter et al. [2013] have created Ultrahaptics (Fig. 14) which is a system that employs focused ultrasound to provide vibratory feedback above the surface. The idea is based on the phenomenon of acoustic radiation force in which ultrasound focused on the skin induces a shear wave that stimulates mechanoreceptors. The system consists of a transducer array, acoustically transparent display above it on which visual content is projected from above, Leap Motion controller for hand tracking and a driver circuit. The transducer array consists of 320 transducers arranged in 16 x 20 formation. Amplitude and phase for each transducer is computed by changing the modulation frequency of the emitted ultrasound and the frequency of the vibration. This way different tactile properties can be attached to a single focal point and versatile feedback can be given using multiple simultaneous feedback points.

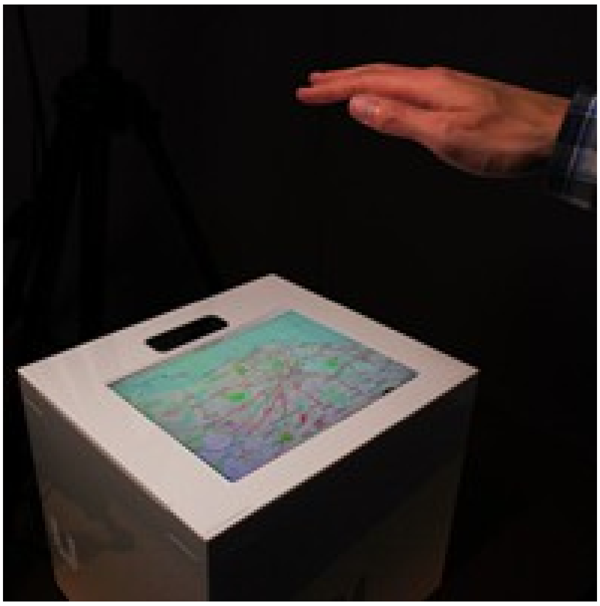


Fig. 14. Ultrahaptics. [Image source: http://big.cs.bris.ac.uk/projects/ultrahaptics]

In their user studies, participants were able to perceive two focal points better when different modulation frequencies were used. Recognition rate of 80% and above were achieved at a separation distance of 3 cm or larger. When the distance was smaller or the same modulation frequencies were used, recognition rates were considerably lower. With training recognition became more accurate.

Obvious advantage of this technique is that no wearable attachments are required. This makes the system easily accessible since users can walk up and start using the device immediately. In Ultrahaptics system feedback can be received from a distance of about 30 cm which is far less than what AIREAL is capable of. The generated feedback, however, is more accurate in Ultrahaptics since it can be targeted at a fingertip whereas air vortex rings can be effectively directed at larger areas.

3.5 Summary

The chapter was divided into three parts. In the first part haptics was briefly defined and functions of mechanoreceptors in producing tactile sensation were explained. In general, haptic feedback is divided into kinesthetic and tactile modalities. In this thesis, the focus is on vibrotactile feedback and other forms of feedback such as thermal or force feedback are beyond the scope of this work. Moreover, the impact of active feedback is the subject of interest in this work because passive feedback is absent in freehand gesture interfaces.

The impact of active feedback in user performance and subjective enjoyability were studied in the second part of the chapter. Study results mostly confirm a benefit of tactile feedback in task performance although contradictory results have also been found. When an aural modality is added as an alternative feedback form, the advantages of vibrotactility are not that clear. It seems that tactile feedback enhances high workload multitask performance. Also, it is beneficial for alarms and warnings.

In non-visual interaction vibration has proven to be effective in informing users of incorrect actions instead of just confirming successful ones. Investigation of tactons has revealed that conveying relatively complex messages through tactile feedback is also possible. Haptic feedback is also potential in providing directional cues and guiding the user's movements which could aid in executing gesture commands in free air.

Whether users accept or dislike vibration seems to depend on the context. Vibration might feel slightly annoying or it can arouse the feeling of being pressured but in environments where visual or aural feedback is not as efficient, tactile feedback is almost consistently preferred. The chosen technology can also have an impact on user preferences.

In the last part, tactile technologies and techniques for touch-based and freehand interaction were presented. As examples of freehand tactile feedback techniques, two novel solutions were presented. AIREAL is a system that utilizes vortices to provide feedback from a distance. Ultrahaptics uses ultrasound to produce localized tactile sensations on the user's hand. In the experiment of this thesis, vibratory stimulation was applied to a finger tip using linear resonant actuators. This technology was chosen because it is a cost-efficient solution and the required equipment is easily available.

4 Experiment

An experiment was conducted to compare two freehand input methods and the effects of visual and vibrotactile feedback on user performance. More precisely, the goal of the experiment was to find out if the two input methods (screentap and pointing) differ in terms of task completion time, characters entered per second, keystrokes entered per second or the number of keystrokes needed to enter the correct character.

It has been shown that vibrotactile feedback can enhance task performance compared to visual modality. Therefore, the aim of the experiment was to investigate if vibrotactile feedback alone improves performance compared to visual-only condition.

4.1 Participants

12 volunteers (mean = 26 years, SD = 8.9 years, range 20-51 years) participated in the experiment. Six of the participants were male. All the participants were right-handed and they had either normal or corrected to normal vision. One participant had abnormal tactile sense in a non-dominant hand but all of the volunteers had normal sensation in their dominant hands. Two of the participants had previous experience of gaming consoles such as Nintendo Wii. All but one were undergraduate students in the University of Tampere. Those volunteers who were students received course credit as a compensation for the participation. One participant terminated the experiment and the data was excluded from the statistical analysis.

4.2 Apparatus

The equipment used in the experiment is presented in Fig. 15. Leap Motion controller (https://www.leapmotion.com/) was used to track positions of the hand and fingers. Leap Motion controller is a 3D tracking device (height: 1.27 cm; width: 3.05 cm; depth: 7.62 cm; weight: 45.36 g) that is able to track all ten fingers up to 1/100th of a millimeter with 150° field of view and track movements at a rate of over 200 frames per second. The controller was connected to Acer Aspire E1-571 laptop (Intel Core™ i7 2.2GHz processor, Intel HD Graphics 4000 graphics card) via USB.

The experimental task was written in Java programming language. The program collected tracking data from Leap Motion and handled also the collection of experimental measurements. Pure Data audio synthesizer software (PD, http://puredata.info/) read the input data from Java program and generated signals for the vibrotactile actuator. One Minebea Linear Vibration Motor actuator (Minebea Matsushita Motor Corporation, LVM-8) was attached to the participant's index finger with electric tape (depicted in Fig. 16). The actuator was attached to the finger throughout the whole experiment. The diameter of an actuator is 0.8 cm and it weights 1.1 g. The actuator was connected to a stereo headphone output of an external USB sound card (ESI Gigaport HD) with a 3.5 mm stereo plug.

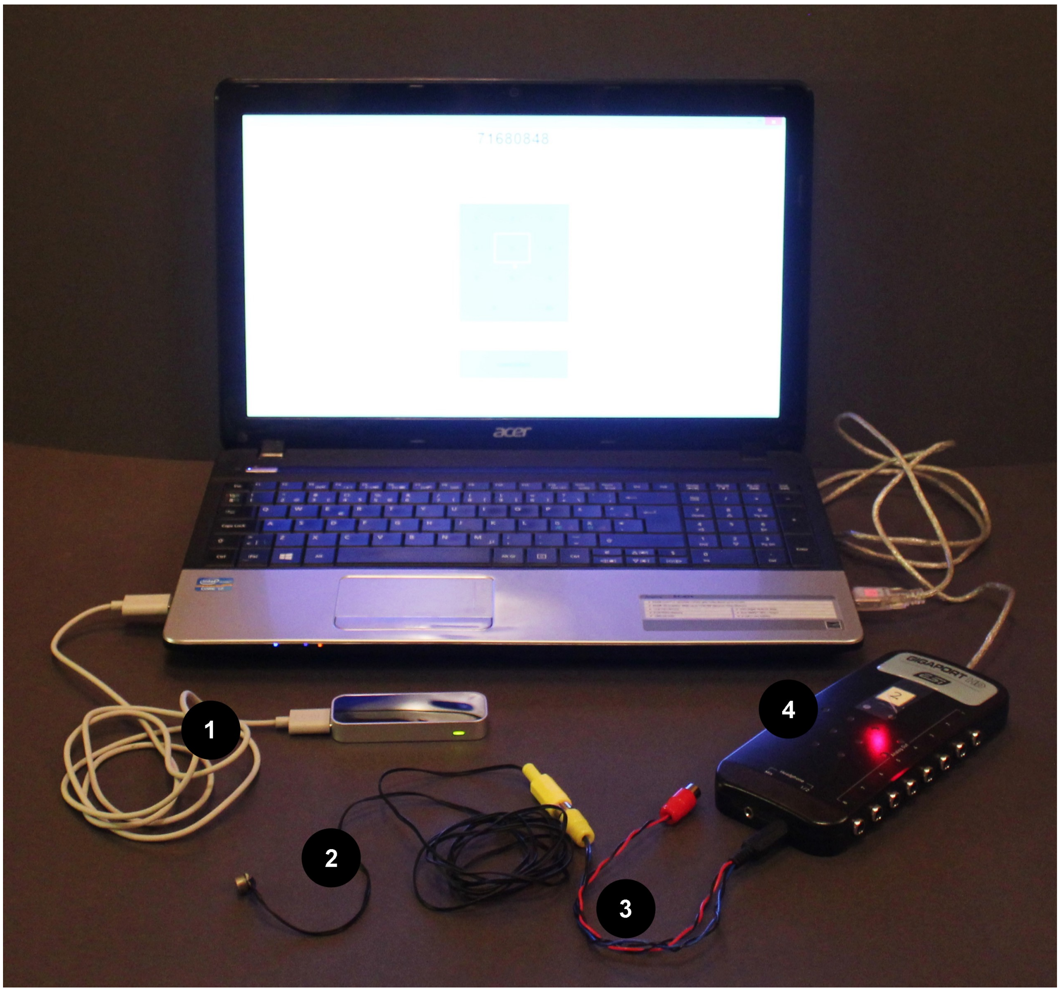


Fig. 15. Experimental equipment. 1) Leap Motion controller connected to a laptop via USB. 2) LVM-8 actuator. 3) The actuator is connected to a sound card with a 3.5 mm stereo plug. 4) ESI Gigaport HD external USB sound card connected to a laptop.

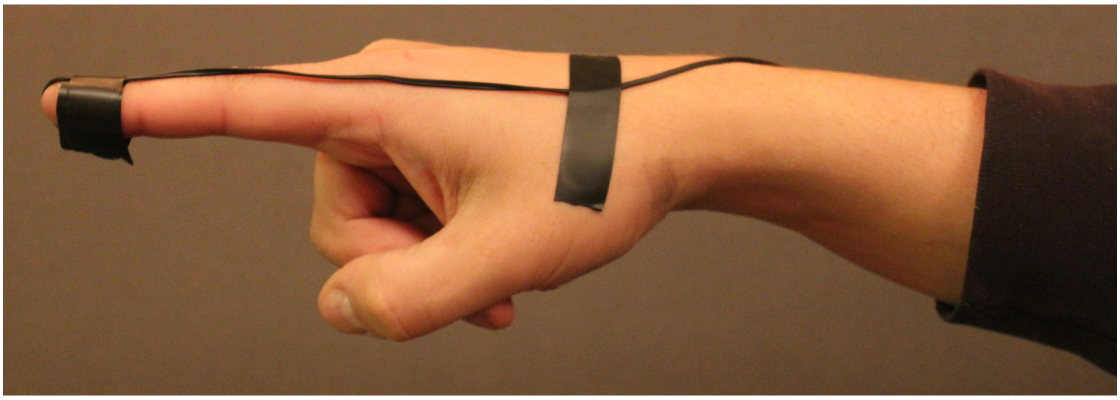
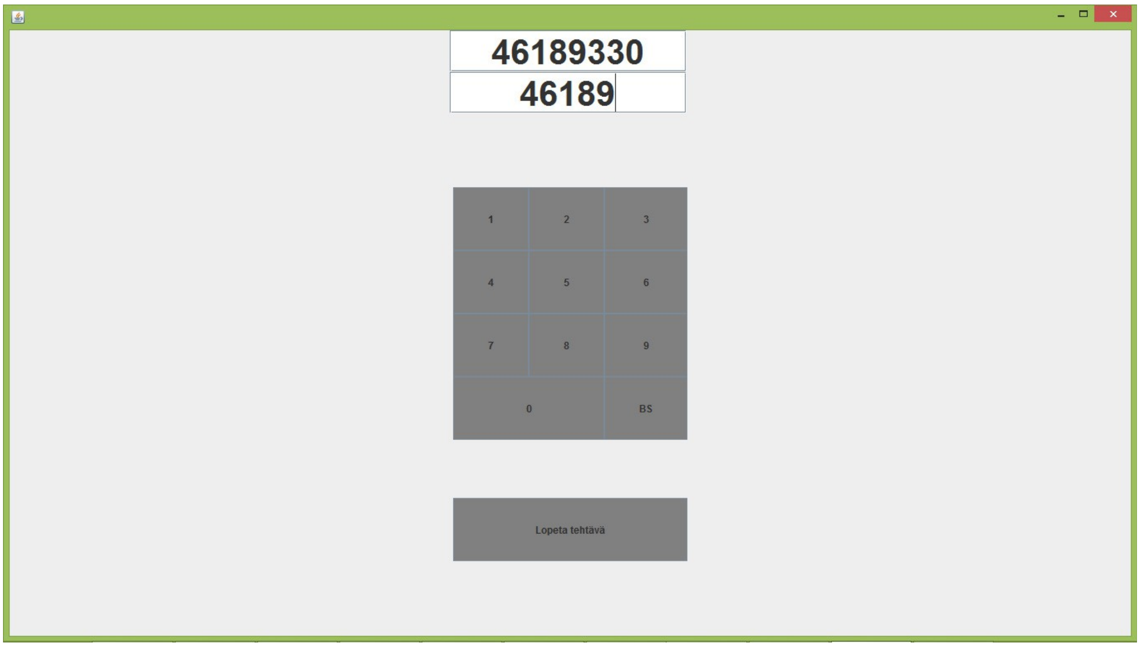


Fig. 16. The hand pose which participants were instructed to hold during the experiment. The actuator and the wire are attached to the participant's hand with electric tape.

4.3 *Experimental application*

A virtual numeric pad (shown in Fig. 17) was implemented to compare two input methods and study the effects of vibrotactile and visual feedback types on task performance. The program was written in Java. Randomly generated number sequence is shown in the upper text field. Digits entered by the participant are shown in the text field below. The numeric pad that resembled one on the traditional keyboard is located in the center of the screen. The numeric pad consists of buttons from 0 to 9 and a backspace button (abbreviation 'BS' used in the label) that is added on the lower right corner. To finish the task, participants clicked the button below the numeric pad. Buttons were 50 x 50 pixels in size except for '0'-button *(100 x 50 pixels)* and 'Finish task'-button *(150 x 50 pixels)*.



46189330

46189

Fig. 17. Virtual numeric pad.

4.4 *Experimental task and stimuli*

The task was to input number sequences that were eight digits long. These number sequences were generated randomly. Tasks were performed using two input gestures. *Screentap* is a forward tapping gesture with a finger. It is performed by tapping as if touching an invisible screen. The gesture is available in Leap Motion SDK by default. *Pointing* method was developed for this experiment. Instead of performing a tapping movement in the air, the participant moves his arm back and forth along the z-axis. Coordinates were normalized to the range [0...1] and the button was clicked when a finger reached the point where z-value was zero (closest to the screen).

For both of these input methods related visual and vibrotactile feedback types were created. In screentap method feedback was linked to hover and mousepress events. Button states during the events are shown in Fig. 18. While cursor hovered over the button, visual feedback highlighted the edges of a button. When the screentap gesture was performed, background colour of a button changed to black for 200 ms. For the creation of vibrotactile feedback, Pure Data generated sine wave with a frequency of 160 Hz. Distinction between hover and mousepress conditions was made by changing the amplitude of a waveform. When the cursor moved over a button, actuator vibrated for 200 ms with an amplitude of 0.25. During the click of a button, actuator vibrated for 200 ms with an amplitude of 1.0 creating a stronger sensation.

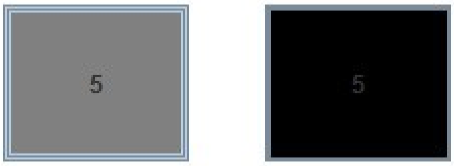


Fig. 18. Button states for the hover and mousepress events when the screentap input method is being used. On the left, visual feedback when cursor is hovering on the button. On the right, visual feedback when the button is clicked.

While using the pointing method, participants received continuous feedback. Visual feedback was implemented by continuously changing sRGB values which were mapped to the fingertip's position on the z-axis. Colour changes are illustrated in Fig. 19. By default button was coloured grey (z = 0.5). When the participant's arm moved closer to the screen, the background colour of a button changed darker and lighter while moving away from the screen. When the finger was closest to the screen, the button was black and white when furthest from the screen.

Generation of continuous vibrotactile feedback was accomplished by changing the amplitude of a sinusoidal waveform that was mapped to the normalized values on the z-axis. Because the linear increase of an amplitude did not produce change in the feedback that was noticeable enough, amplitude was increased exponentially.

Fig. 19. From left to right: button colour when z-value is 1, 0.75, 0.5, 0.25 and 0.

4.5 *Procedure*

A participant was seated in front of the computer and the moderator shortly explained the experimental task and presented the devices used in the experiment. The moderator also demonstrated in the air how to perform both of the input methods. Participants were instructed to keep their index finger straight and other fingers clenched in a fist to ensure accurate localisation of a finger tip (Fig. 16). Each participant was adviced to adjust their arm position during the tasks in order to ensure correct and stable hand tracking. If the participant was struggling with gesturing during the tasks, the moderator only reminded about the adjustment of the arm or finger position. Vibrotactile and visual feedback types were described in detail to the participant. Before the test, participants were asked to sign a written consent and they also filled a background questionnaire.

The actual task procedure began with the moderator preparing a series of tasks on a computer. When the program was started, 'Start' button appeared on the screen (Fig. 20). The participant clicked this button using the input method that was to be used during the tasks. After clicking the button, a short text informing about the start of the task appeared on the screen for ten seconds. A counter showed how many seconds there were left before the beginning of an upcoming task. The text and the counter were shown before every individual task. When the time was up, the numeric pad appeared on the screen. Participants did not have to start entering digits immediately but they were instructed to perform each task as quickly as possible and without interruptions after they had entered the first digit. Furthermore, participants were asked to make exact copies of the number sequences. The task was finished by pressing the 'Finish task' button below the numeric pad. If the participant wanted to skip the task, this was done by pressing the space bar on the keyboard.

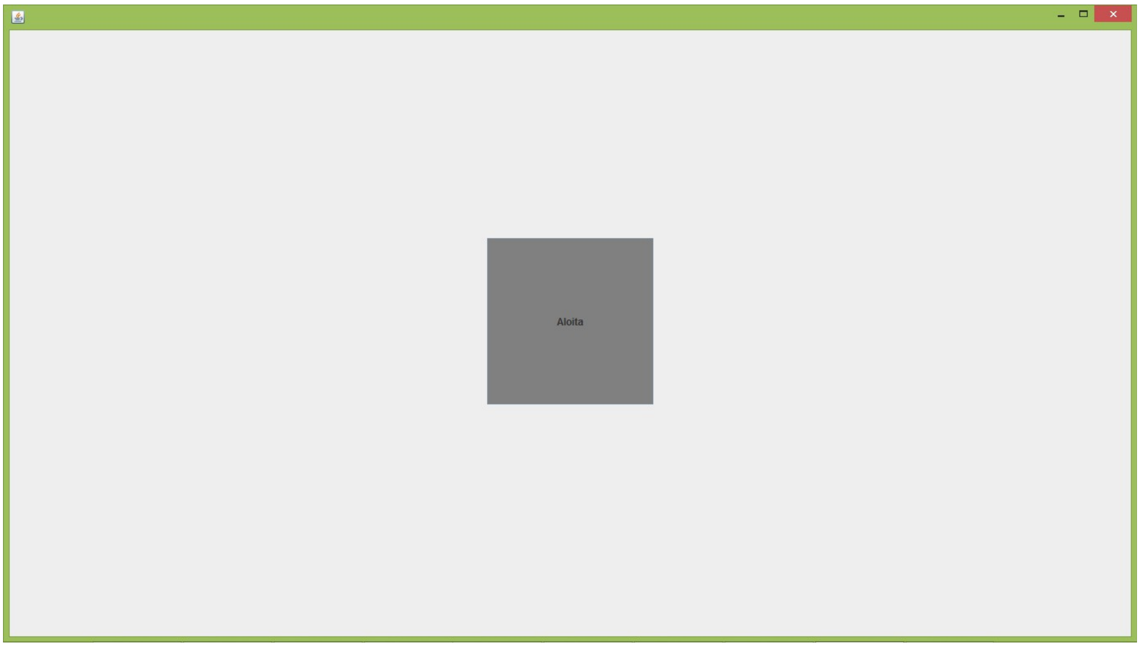


Fig. 20. When the Start-button appeared on the screen, control was passed to the participant. Button is 200 x 200 (px) in size.

Participants performed one practice task and five actual tasks in every condition. After finishing all the five tasks, the participant evaluated the particular method by filling the NASA Task Load Index (NASA-TLX) questionnaire [Hart and Staveland, 1988]. The whole procedure was repeated four times, each time with a different input or feedback method. The order of the screentap and feedback combinations was counterbalanced using Latin square. Finally, participants were asked to rank the interaction styles (1 = the best, 4 = the worst) and write subjective comments about what was good or bad with the input and feedback methods.

4.6 Experiment design and data analysis

The experiment used a 2 x 2 within-factor design. The independent variables were feedback (vibrotactile and visual) and input method (screentap and pointing). Dependent variables in the experiment were task completion time, characters per second (CPS), keystrokes per second (KSPS) and keystrokes per character (KSPC). A two-way repeated measures analysis of variance (ANOVA) with Bonferroni-corrected pairwise comparisons was performed in the quantitative data analysis.

The NASA-TLX questionnaire was used to measure subjective workload. The questionnaire consists of six component scales which are mental demand, physical demand, temporal demand, performance, effort and frustration. The questionnaire was translated to Finnish from an English version. A two-way repeated measures ANOVA was used in the analysis of workload data. Participants also ranked interaction styles from best to worst and they were asked to write down subjective comments about the advantages and disadvantages of the input and feedback methods.

4.7 Results

Results of the statistical analysis and subjective evaluations are presented next. First, quantitative measurements including task completion time, characters per second, keystrokes per second and keystrokes per character are reported. Second, data gathered from the NASA-TLX questionnaire and rankings of the interaction methods are analysed.

4.7.1 Quantitative measurements

Mean task completion times and standard error of the means are shown in Fig. 21. A two-way 2 x 2 (input method x feedback) ANOVA on task completion time showed a statistically significant main effect for input method ($F(1, 59) = 6.797$, $p < 0.05$). The main effect of feedback or the interaction of the main effects were not statistically significant. Post-hoc pairwise comparison showed that pointing method was significantly faster than screentap method (*MD* = 19.04, $p < 0.05$).

Mean characters per second and standard error of the means are presented in Fig. 22. A two-way ANOVA on CPS showed a statistically significant interaction effect ($F(1, 59) = 4.421$, $p < 0.05$). The main effects for input method and feedback were not statistically significant. Further analysis using one-way ANOVA on input and feedback methods did not show statistically significant differences between group means.

Mean keystrokes per second and standard error of the means are presented in Fig. 23. A two-way ANOVA on KSPS showed a statistically significant main effect for input method ($F(1, 59) = 5.953$, $p < 0.05$) and a statistically significant interaction effect ($F(1, 59) = 5.063$, $p < 0.05$). The main effect for feedback was not statistically significant. To further analyse the interaction of the main effects, a one-way ANOVA was performed on input method comparing pointing and screentap styles and feedback method comparing visual and tactile modalities. Statistically significant difference was found between group means of input methods ($F(1, 238) = 5.943$, $p < 0.05$) showing that screentap method induced more keystrokes per second than pointing method. No statistical difference was found between group means comparing feedback methods.

Mean keystrokes per character and standard error of the means are presented in Fig. 24. A two-way ANOVA on KSPC showed a statistically significant main effect for input method ($F(1, 59) = 3.991$, $p \leq 0.05$) and a statistically significant main effect for feedback ($F(1, 59) = 4.440$, $p < 0.05$). Post-hoc comparisons showed that participants used less keystrokes per character with pointing method (*MD* = -0.175, $p < 0.05$) and less keystrokes per character with visual feedback (*MD* = -0.154, $p < 0.05$). The interaction of the main effects was not statistically significant.

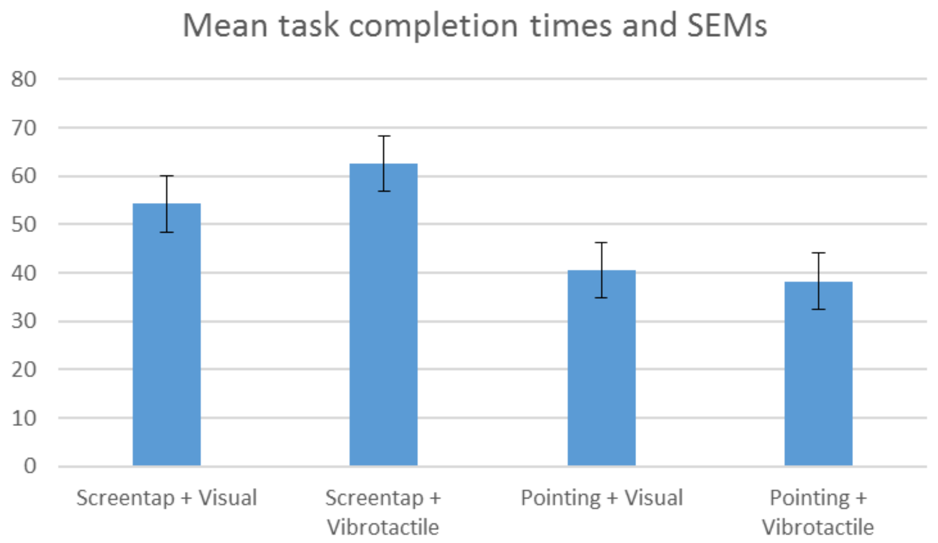


Fig. 21. Mean task completion times in seconds and standard error of the means.

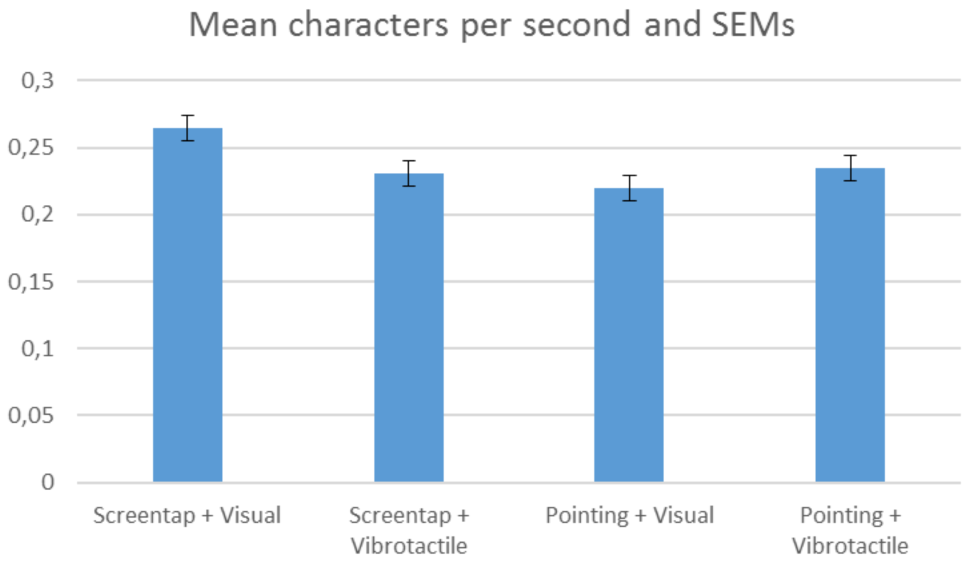


Fig. 22. Mean characters per second and standard error of the means.

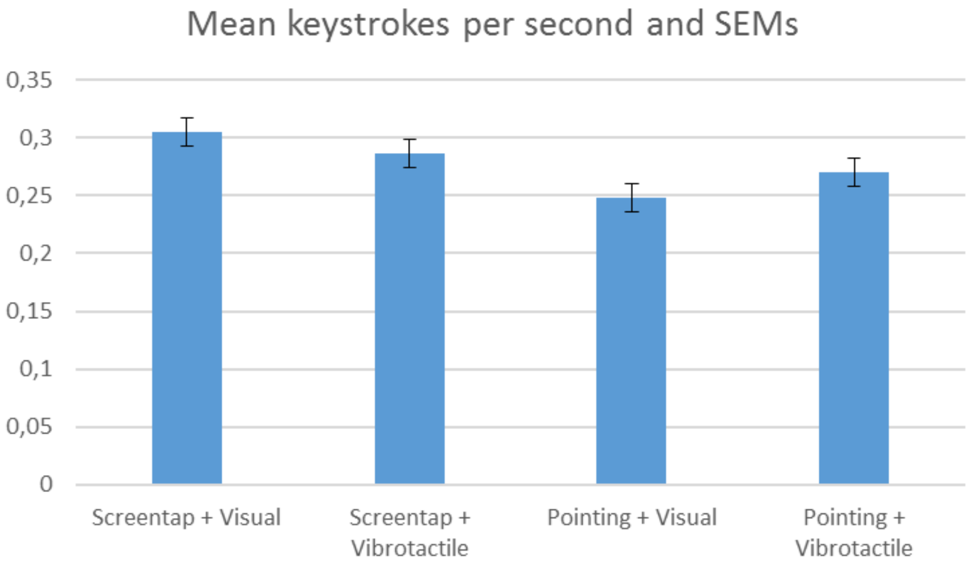


Fig. 23. Mean keystrokes per second and standard error of the means.

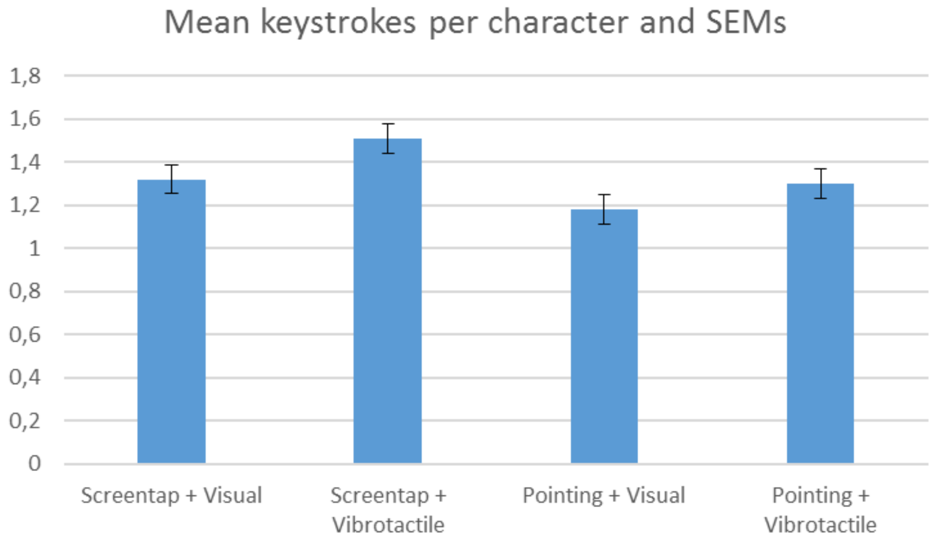


Fig. 24. Mean keystrokes per character and standard error of the means.

4.7.2 Subjective measurements

Mean ratings and standard error of the means are presented in Fig. 25. A two-way ANOVA was performed on each component scale of the NASA-TLX questionnaire. For the ratings of frustration, a two-way ANOVA showed a statistically significant main effect of input method ($F(1, 11) = 12.108$, $p < 0.05$). Post-hoc pairwise comparison showed that pointing was rated significantly less frustrating compared to screentap method ($MD = -3.792$, $p < 0.05$). For other ratings, the main effects or the interaction of the main effects were not statistically significant.

Rankings of the interaction styles are presented in Fig. 26. Regardless of the feedback type pointing as an input method was ranked the highest. Both feedback methods received an equal number of highest rankings (4/12). However, pointing combined with continuous tactile feedback was ranked more often as the second best (5/12) suggesting it was the most preferred alternative. Rankings imply that screentap coupled with vibrotactile feedback was the least preferred method. Seven out of twelve participants ranked it as the worst. However, the same method was rated as the best alternative by three participants whereas only one participant rated the screentap method with visual feedback as the best technique.

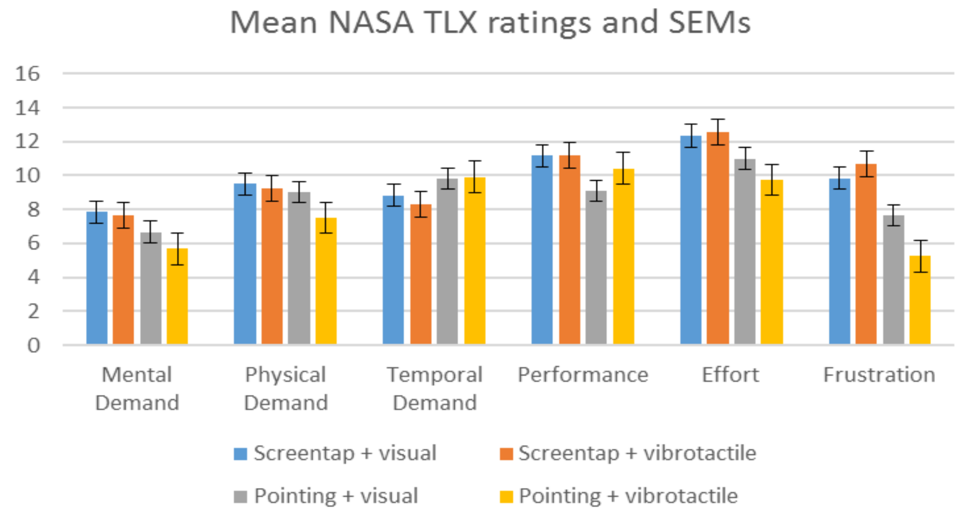


Fig. 25. Results from the NASA TLX questionnaire.

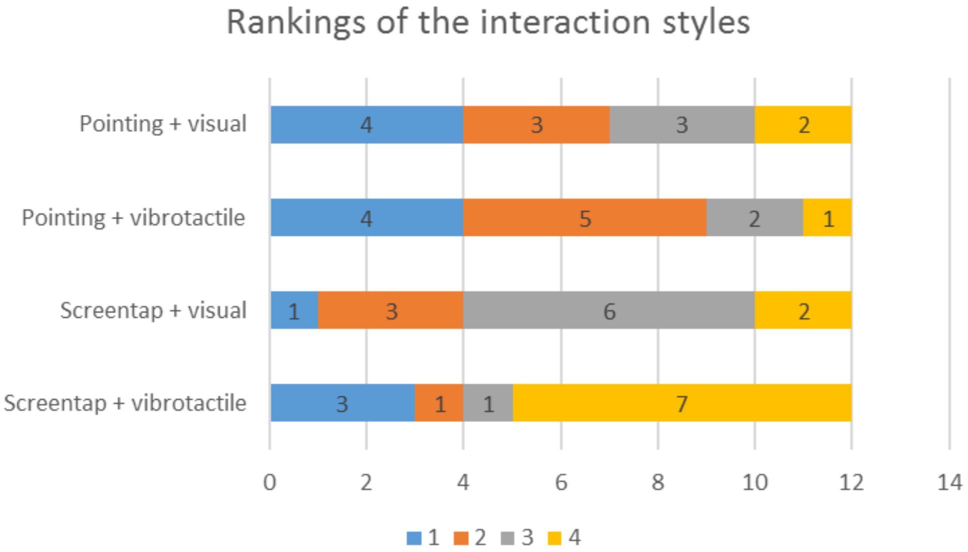


Fig. 26. Results from the rankings of interaction styles. Each level of the scale is colour coded. These are explained on the bottom of the figure (1 = the best … 4 = the worst). Frequencies are shown inside bars.

4.8 *Discussion*

Based on the analysis, it can be concluded that considerably faster performance can be achieved with pointing method. On average, screentap technique was almost 20 seconds slower when typing 80 numbers in total with each input method. Since participants were required to make perfect copies of the presented number sequences, faster performance could be explained by smaller error rate. This appears to be the case since less keystrokes per character were needed when participants used the pointing technique. Using screentap, participants performed more keystrokes per second indicating that the targets were missed more often.

Apparently pointing as an input method is more accurate and faster than screentap gesture for novices as most of the participants were inexperienced users of gesture technology and none of them had used Leap Motion before the experiment. However, screentap would be expected to lead to better performance after practice. With screentap participants tried to hit the buttons more frequently suggesting that this technique would outperform pointing when users become more accurate after training. It would be interesting to find out how long it takes for screentap to become more efficient than pointing.

Due to the poor accuracy of the screentap method participants felt frustrated. Data analysis of subjective workload questionnaire confirms this as screentap method was rated considerably more frustrating than pointing technique. Frustration also led to overshooting and participants forcefully tried to hit the buttons which in turn resulted in performing undetectable gestures. It seems that screentap gesture requires overly precise hand movement that is executed with correct speed and within a certain range of length.

When feedback is considered, results are not in line with previous studies showing the benefits of tactile feedback (e.g. Krol et al. [2009]; Adams et al. [2010]; Lehtinen et al. [2012]). Results are similar to those of Foehrenbach et al. [2009] who did not confirm a benefit of haptic feedback in mid-air gesture interaction. The only significant difference in this experiment was found measuring KSPC and it appears that visual feedback produced more accurate and more efficient performance as less keystrokes were required per character. Measurements of task completion time did not reveal significant difference between modalities. When CPS and KSPS measurements were analysed, results suggested that vibrotactile feedback enhanced performance when combined with a pointing method. A slight increase in the means was observed. On the contrary, the screentap method seems to gain from visual feedback as opposite findings were obtained. However, no definitive conclusions should be made since significant differences between feedback types were not obtained.

Although quantitative measurements cannot fully support the hypothesis for the positive effect of vibrotactile feedback, subjective evaluations show otherwise. Vibrotactile feedback was considered to be more recognizable. One participant commented it was difficult to differentiate between shades of black in continuous visual feedback and it was not always clear at which point the button was clicked. One person also brought up the calming effect of vibrotactile feedback due to the sense of control it provided. Tactile feedback was thought to guide hand movement well. Visual feedback was preferred by some participants who considered tactile feedback to be annoying and visual feedback more informative.

Participants preferred pointing technique because it was more stable and it was easier to target the buttons. Almost every participant commented that screentap was more difficult than pointing method and took longer to learn. It was also hard to estimate the correct speed and distance of motion required to perform the gesture. Only one participant considered the combination of screentap and visual feedback to be the most pleasant and the most natural alternative.

Some technical difficulties occurred during the experiment. There were problems with continuous visual feedback due to unstabilities in hand tracking. Also, some of the participants struggled performing the screentap gesture. It is difficult to say to what extent they were caused by the system being unable to detect hand gestures or participants performing gestures incorrectly.

Obviously, one disadvantage of the tactile feedback method presented here is that the finger needs to be equipped with an external device. Of course, this impairs the accessibility of the interaction. Nonetheless, the goal of the experiment was to study the effects of two feedback methods on performance and not the technology itself. Using a vibration motor actuator is an inexpensive and simple technique to generate haptic feedback. The research in the area of contactless methods can also benefit from the knowledge that is obtained in the experiments studying feedback received through wearable devices.

Finally, a few words about the practicality of freehand data entry. It is clear that it will not replace traditional keyboards or touch-based entry methods. Among all the techniques studied, CPS of 0.62 was the highest measurement. Translated to words per minute, the rate of only seven words per minute could be achieved at best. However, there are a few scenarios that come to mind where this kind of interaction could be appropriate. One is an environment where the user needs to wear bulky gloves that inhibits the ability to use a traditional keyboard or a touchscreen. In a study of Adams et al. [2010] text entry in free air for astronauts was investigated. Options for data entry in an extravehicular environment are limited so hand tracking combined with vibrotactile feedback embedded in a space suit glove is one suitable solution. Also, in sterile environments keyboard and touchscreen usage are restricted. For example, freehand gesture interaction can be beneficial in surgical operating rooms. Of course, tactile feedback would also have to be contactless. Concerning sterility, freehand tapping or pointing could also be a pleasant option for interaction with public displays such as ATMs or information screens.

5 Summary

This thesis work investigated vibrotactile feedback in freehand gesture interaction. The primary goal of this work was to examine how vibrotactile feedback could enhance gestural interaction. The effects on task performance and user satisfaction were investigated. Another objective was related to the design of gesture interaction. The properties of a single gesture command and the learnability of freehand interaction were studied.

An experiment was carried out to compare two input methods and to find out whether vibrotactile feedback enhances performance in a freehand data entry task when compared to visual-only condition. The findings indicate that the input method developed for the purposes of the experiment produced significantly faster and more accurate performance than Leap Motion's default screentap gesture. The pointing method was also considered significantly less frustrating. Regarding the feedback methods, the only significant difference was found measuring the number of keystrokes required to enter a character. Subjective evaluations suggested that vibrotactile feedback was considered to be more precise and recognizable than visual feedback.

Even though there are still difficulties related to the design of gestural input and the implementation of contactless tactile feedback, the future looks bright for freehand gesture interfaces and haptic feedback technology. In the recent years freehand gestural interfaces have been increasingly spreading into the everyday lives of users. Products like Xbox Kinect have already familiarised people with this type of interaction. It has been one of the most fastest selling products and already tens of millions of copies have been sold since its launch in 2010. Interactive TVs are also finding their way to the living rooms of consumers.

Today it is also easier to create gesture-based applications since sensor technology has become affordable and software development kits for individual working are provided. Knowledge of recognition techniques is not necessarily required and developers can concentrate on the design of interaction. However, the easiness of implementation could also be a stumbling block for developers because the challenging task actually is to redesign the already established interaction methods or entirely replace them with new ideas.

In the area of haptics more and more intriguing and innovative solutions can be anticipated in the near future since consumer electronics sector has started to invest strongly in novel technology. Haptics market is expected to grow rapidly in the next decade. Lux Research Inc. (http://www.luxresearchinc.com/) has forecasted the market to reach 9.8 billion USD which is around 11 times today's value. According to predictions touch technology solutions will prevail and public interfaces will form the second largest market.

Perhaps haptic feedback and gestural input methods could be developed for public interfaces. ATMs or information displays could be controlled with gestures and ultrasonic transducers, for instance, could be used to provide tactile sensations on top of a surface.

One interesting area of research has been the study of haptic passwords. The fundamental idea is to use personalized tactons instead of PINs. This kind of solution could improve privacy of the users and remove the possibility to steal passwords by peeking over the shoulders. Right now, technology is touch-based but it would be interesting to find out whether tactons can be reliably recognized in free air.

The experiment conducted for this thesis also involved interaction that is close to above-surface style interaction. The results, nonetheless, do not favor freehand input method for data entry tasks. Although the design of an alternative input method was successful, based on the results this kind of interaction may never outperform direct control in speed and accuracy even though hand and finger tracking would be perfectly reliable and stable. The problem is that the tapping and pointing methods try to mimic actions similar to pushing the real buttons. In most situations it would be more suitable to just use a keyboard, either real or a virtual one. For situations where this is not possible, instead of performing keyboard control in free air, interaction should be designed differently.

Perhaps letters and digits could be drawn in free air. Based on the results of elicitation studies, this kind of gesturing is intuitive for users. At the same time, the complexity of the implementation is increased due to the diversity of possible patterns for the same symbol. It should also be considered if mid-air gestural input is appropriate for above-surface interaction or performed near the screen. Perhaps gesturing is truly advantageous when the interaction takes place far away from the display. Tactile cueing and haptic guidance techniques could significantly improve user performance. Gesturing in front of a large screen would also make the group work possible.

Social aspect has been repeatedly mentioned in the discussion of natural user interfaces. When looking at the issue from a learning perspective, enabling simultaneous actions of multiple users would strengthen the feeling of naturalness in interaction. When users are working as a group, they can explore all the possible actions together and learn by observing, copying and teaching each other. Educational and work settings would especially benefit from cooperative multi-user applications.

It will be fascinating to see in what direction the development of haptic and gesture technologies will go in the coming years. If predictions of the expanding market prove to be correct, the feedback arsenal will be strengthened with one essential information channel.

References

[Adams et al., 2010] Richard J. Adams, Aaron B. Olowin, Blake Hannaford, and O. Scott Sands, Tactile Data Entry for Extravehicular Activity. In: *2011 IEEE World Haptics Conference (WHC)*, 305-310.

[Aigner et al., 2012] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, J. T. K. V. Koh, Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Microsoft Research TechReport **MSR-TR-2012-111.**

[Ashbrook and Starner, 2010] D. Ashbrook and T. Starner, MAGIC: a motion gesture design tool. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 2159-2168.

[Blackler and Hurtienne, 2007] Alethea Liane Blackler and Jorn Hurtienne, Towards a unified view of intuitive interaction: definitions, models and tools across the world. *MMI-interaktiv,* **13,** 36-54.

[Bau et al., 2010] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, TeslaTouch: electrovibration for touch surfaces. In: *Proc. of the 23nd annual ACM symposium on User interface software and technology,* 283-292.

[Bobeth et al., 2014] J. Bobeth, J. Schrammel, S. Deutsch, M. Klein, M. Drobics, C. Hochleitner, and M. Tscheligi, Tablet, gestures, remote control?: influence of age on performance and user experience with iTV applications. In: *Proc. of the 2014 ACM international conference on Interactive experiences for TV and online video,* 139-146.

[Bolanowski et al., 1988] S. J. Bolanowski, G. A. Gescheider, R. T. Verrillo, and C.M. Checkosky, Four channels mediate the mechanical aspects of touch. *J. Acoust. Soc. Am.*, **84,** 5 (1988), 1680-1694.

[Brewster et al., 2007] S. Brewster, F. Chohan, L. Brown, Tactile feedback for mobile interactions. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07),* 159–162.

[Brown et al., 2005] L. M. Brown, S. A. Brewster, and H. C. Purchase, A first investigation into the effectiveness of tactons. In: *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint,* 167-176.

[Burke et al., 2006] Jennifer L. Burke, Matthew S. Prewett, Ashley A. Gray, Liuquin Yang, Frederick R. B. Stilson, Michael D. Coovert, Linda R. Elliot, Elizabeth Redden, Comparing the Effects of Visual-Auditory and Visual-Tactile Feedback on User Performance: A Meta-analysis. In: *Proc. of the 8th international conference on Multimodal interfaces (ICMI '06)*, 108-117.

[Cadoz, 1994] Claude Cadoz, *Les réalités virtuelles*. Flammarion, 1994.

[Carter et al., 2013] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian, UltraHaptics: Multi-Point Mid-Air Haptic Feedback for Touch Surfaces. In: *Proc. of the 26th annual ACM symposium on User interface software and technology (UIST '13),* 505-514.

[Charoenchaimonkon et al., 2010] E. Charoenchaimonkon, P. Janecek, M. N. Dailey, and A. Suchato, A Comparison of Audio and Tactile Displays for Non-Visual Target Selection Tasks. In: *2010 International Conference on User Science and Engineering (i-USEr),* 238-243.

[Cox et al., 2012] D. Cox, J. Wolford, C. Jensen, and D. Beardsley, An evaluation of game controllers and tablets as controllers for interactive tv applications. In: *Proc. of the 14th ACM international conference on Multimodal interaction,* 181-188.

[Dennerlein et al., 2000] Jack Dennerlein, David Martin, and Christopher Hasser, Force-feedback improves performance for steering and combined steering-targeting tasks. *CHI Letters,* **2**, 1, 423-429.

[Efron, 1941] David Efron, *Gesture and Environment*. Morningside Heights, New York: King's Crown Press.

[Foehrenbach et al., 2009] Stephanie Foehrenbach, Werner A. König, Jens Gerken, and Harald Reiterer, Tactile feedback enhanced hand gesture interaction at large, high-resolution displays, *Journal of Visual Languages & Computing,* **20**, 5 (October 2009), 341-351.

[Gescheider et al., 2002] George A. Gescheider, Stanley J. Bolanowski, Jennifer V. Pope, and Ronald T. Verrillo, A four-channel analysis of the tactile sensitivity of the fingertip: frequency selectivity, spatial summation, and temporal summation. *Somatosens. Mot. Res.*, **19,** 2 (2002), 114-124.

[Hachisu and Fukumoto, 2014] T. Hachisu and M. Fukumoto, VacuumTouch: attractive force feedback interface for haptic interactive surface using air suction. In: *Proc. of the 32nd annual ACM conference on Human factors in computing systems,* 411-420.

[Hart and Staveland, 1988] S. G. Hart and L. E. Staveland, Development of Nasa-Tlx (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload* (1988), 139-183.

[Heidrich et al., 2011] F. Heidrich, M. Ziefle, C. Röcker, and J. Borchers, Interacting with smart walls: a multi-dimensional analysis of input technologies for augmented environments. In: *Proc. of the 2nd Augmented Human International Conference,* 1-8.

[Hincapié-Ramos et al., 2014] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. In: *Proc. of the 32nd annual ACM conference on Human factors in computing systems,* 1063-1072.

[Hoggan et al., 2008] Eve Hoggan, Stephen A. Brewster, and Jody Johnston, Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 1573-1582.

[Hoggan et al., 2009] Eve Hoggan, Andrew Crossan, Stephen Brewster, and Topi Kaaresoja, Audio or Tactile Feedback: Which Modality When? In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 2253-2256.

[Jacko et al., 2003] Julie Jacko, Ingrid Scott, Francois Sainfort, Leon Barnard, Paula Edwards, Kathlene Emery, Thitima Kongnakorn, Kevin Moloney, and Brynley Zorich, Older adults and visual impairment: what do exposure times and accuracy tell us about performance gains associated with multimodal feedback? *CHI Letters,* **5,** 1, 33-40.

[Johnson, 2001] Kenneth O. Johnson, The roles and functions of cutaneous mechanoreceptors. *Curr. Opin. Neurobiol.*, **11,** 4, 455-461.

[Karam and Schraefel, 2005] Maria Karam and M.C. Schraefel, A taxonomy of gestures in human computer interaction. University of Southampton, Electronics and Computer Science, Technical report **ECSTR-IAM05-009**.

[Kendon, 1988] Adam Kendon, How gestures can become like words. In: F. Poyatos (ed.), *Crosscultural perspectives in nonverbal communication*. Hogrefe, 1988, 131-141.

[Kendon, 2004] Adam Kendon, *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

[Koskinen et al., 2008] Emilia Koskinen, Topi Kaaresoja and Pauli Laitinen, Feel-Good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button. In: *Proc. of the 10th international conference on Multimodal interfaces (CHI '08)*, 297-304.

[Krol et al., 2009] Laurens R. Krol, Dzmitry Aliakseyeu, and Sriram Subramanian, Haptic Feedback in Remote Pointing. In: *Proc. of CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09),* 3763-3768.

[Kurtenbach, 1993] G. P. Kurtenbach, *The design and evaluation of marking menus.* Doctoral dissertation, University of Toronto.

[Lee et al., 2004] Lee, J. C., Dietz, P. H., Leigh, D., Yerazunis, W. S., and Hudson, S. E., Haptic pen: a tactile feedback stylus for touch screens. In: *Proc. of the 17th annual ACM symposium on User interface software and technology,* 291-294.

[Lehtinen et al., 2012] V. Lehtinen, A. Oulasvirta, A. Salovaara, and P. Nurmi, Dynamic tactile guidance for visual search tasks. In: *Proc. of the 25th annual ACM symposium on User interface software and technology,* 445-452.

[Maike et al., 2014] V. R. M. L. Maike, L. D. S. B. Neto, M. C. C. Baranauskas, and S. K. Goldenstein, Seeing through the Kinect: A Survey on Heuristics for Building Natural User Interfaces Environments. In: *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access, Lecture Notes in Computer Science,* **8513,** 407-418, Springer International Publishing.

[Martin and Parikh, 2011] Michael W. Martin and Sarangi P. Parikh, Improving Mobile Robot Control - Negative Feedback for Touch Interfaces. In: *2011 IEEE Conference on Technologies for Practical Robot Applications (TePRA),* 70-75.

[McNeill, 1992] David McNeill, *Hand and Mind: What Gestures Reveal About Tought*. Chicago: The university of Chicago Press, 1992.

[McNeill, 2006] David McNeill, Gesture and communication. In: Keith Brown (ed.), *Encyclopedia of Language & Linguistics (Second Edition)*, 2006, 58-66.

[Mitra and Acharya, 2007] S. Mitra and T. Acharya. Gesture recognition: a survey. *IEEE Trans. Syst. Man. Cybern. C Appl. Rev.,* **37**, 3, 311-324.

[Mulder, 1996] A. Mulder, Hand gestures for HCI. Hand centered studies of human movement project. Simon Fraser University, School of Kinesiology, Technical report **96-1**. Also available: http://xspasm.com/x/sfu/vmi/HCI-gestures.htm

[Nancel et al., 2011] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay, Mid-air pan-and-zoom on wall-sized displays. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 177-186.

[Nielsen, 1994] J. Nielsen, Heuristic evaluation. In J. Nielsen, and R. L. Mack (eds.), *Usability Inspection Methods.* John Wiley & Sons, New York, NY.

[Norman, 2004] Donald Norman, *Affordances and design*. http://www.jnd.org/dn.mss/affordances_and_desi.html Checked 8.10.2014.

[Norman, 2010] Donald A. Norman, Natural user interfaces are not natural. *Interactions,* **17**, 3 (May – June 2010), 6-10.

[O'Hara et al., 2013] Kenton O'Hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor, On the naturalness of touchless: putting the "interaction" back into NUI. *ACM Trans. Comput. Hum. Interact. (TOCHI) - Special issue on the theory and practice of embodied interaction in HCI and interaction design,* **20**, 1 (March 2013), article no. 5.

[Pasquero and Hayward, 2011] J. Pasquero and V. Hayward, Tactile feedback can assist vision during mobile interactions. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 3277-3280.

[Pfeiffer et al., 2014] Max Pfeiffer, Stefan Schneegass, Florian Alt, and Michael Rohs, Let Me Grab This: A Comparison of EMS and Vibration for Haptic Feedback in Free-Hand Interaction. In: *Proc. of the 5th Augmented Human International Conference (AH '14)*, article no. 48.

[Pyryeskin et al., 2012] D. Pyryeskin, M. Hancock, and J. Hoey, Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In: *Proc. of the 2012 ACM international conference on Interactive tabletops and surfaces,* 1-10.

[Quek et al., 2002] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari, Multimodal human discourse: gesture and speech. *ACM T. Comput.-Hum. Int.,* **9**, 3 (Sept. 2002), 171-193.

[Rovan and Hayward, 2000] J. Rovan and V. Hayward. Typology of tactile sounds and their synthesis in gesture-driven computer music performance. *Trends in gestural control of music*, 297-320.

[Ruiz et al., 2011] Jaime Ruiz, Yang Li, and Edward Lank, User-Defined Motion Gestures for Mobile Interaction. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 197-206.

[Sodhi et al., 2012] R. Sodhi, H. Benko, and A. Wilson, LightGuide: projected visualizations for hand movement guidance. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 179-188.

[Sodhi et al., 2013] Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr, AIREAL: interactive tactile experiences in free air. *ACM Trans. Graph. (TOG) - SIGGRAPH 2013 Conference Proceedings*, **32**, 4 (July 2013), article no. 134.

[Songh et al., 2012] P. Song, W. B. Goh, W. Hutama, C. W. Fu, and X. Liu, A handle bar metaphor for virtual object manipulation with mid-air interaction. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 1297-1306.

[Subramanian et al., 2005] S. Subramanian, C. Gutwin, M. Nacenta, C. Power, and L. Jun, Haptic and Tactile Feedback in Directed Movements. In: *Proc. of conference on Guidelines on Tactile and Haptic Interactions*, 2005.

[Vatavu, 2012] R. D. Vatavu, User-defined gestures for free-hand TV control. In: *Proc. of the 10th European conference on Interactive tv and video,* 45-48.

[Vatavu and Zaiti, 2014] R. D. Vatavu and I. A. Zaiti, Leap gestures for TV: insights from an elicitation study. In: *Proc. of the 2014 ACM international conference on Interactive experiences for TV and online video,* 131-138.

[Wachs et al., 2008] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith, A gesture-based tool for sterile browsing of radiology images. *J. Am. Med. Inform. Assoc.*, **15**, 3, 321-323.

[Walter et al., 2013] R. Walter, G. Bailly, and J. Müller, Strikeapose: revealing mid-air gestures on public displays. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems,* 841-850.

[Weber et al., 2011] B. Weber, S. Schatzle, T. Hulin, C. Preusche, and B. Deml, Evaluation of a vibrotactile feedback device for spatial guidance. In: *2011 IEEE World Haptics Conference (WHC),* 349-354.

[Wigdor and Wixon, 2011] D. Wigdor and D. Wixon, *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann, 2011.

[Wobbrock et al., 2009] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson, User-Defined Gestures for Surface Computing. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 1083-1092.

[Wu et al., 2006] J.Z. Wu, K. Krajnak, D.E. Welcome, and R.G. Dong, Analysis of the dynamic strains in a fingertip exposed to vibrations: correlation to the mechanical stimuli on mechanoreceptors. *J. Biomech.*, **39**, 13 (2006), 2445-2456.

[Wu and Wang, 2013] H. Wu and J. Wang, Understanding user preferences for freehand gestures in the TV viewing environment. *AISS*, **5**, 709-717.

[Zamborlin et al., 2014] B. Zamborlin, F. Bevilacqua, M. Gillies, and M. d'Inverno, Fluid gesture interaction design: applications of continuous recognition for the design of modern gestural interfaces. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, **3,** 4, 22.