# Understanding 3D Mid-Air Hand Gestures with Interactive Surfaces and Displays: A Systematic Literature Review

Celeste Groenewald, Craig Anslow, Junayed Islam, Chris Rooney, Peter Passmore, William Wong
Department of Computer Science
Middlesex University
London, NW4 4BT, UK
{*c.groenewald,c.anslow,j.islam,c.rooney,p.passmore,w.wong*}*@mdx.ac.uk*

**3D gesture based systems are becoming ubiquitous and there are many mid-air hand gestures that exist for interacting with digital surfaces and displays. There is no well defined gesture set for 3D mid-air hand gestures which makes it difficult to develop applications that have consistent gestures. To understand what gestures exist we conducted the first comprehensive systematic literature review on mid-air hand gestures following existing research methods. The results of the review identified 65 papers where the mid-air hand gestures supported tasks for selection, navigation, and manipulation. We also classified the gestures according to a gesture classification scheme and identified how these gestures have been empirically evaluated. The results of the review provide a richer understanding of what mid-air hand gestures have been designed, implemented, and evaluated in the literature which can help developers design better user experiences for digital interactive surfaces and displays.**

*3D Gestures, HCI, Interaction Gestures, Mid-Air Hand Gestures, Systematic Literature Review.*

## 1. INTRODUCTION

3D tracking systems in real time allow detecting gestures by humans via their hands and full body to interact with applications Bowman et al. (2004). Gesture-based systems are now becoming ubiquitous with devices such as the Microsoft Kinect and Leap Motion which support tracking of hands and full body. With the advent of mixed reality systems such as Microsoft Hololens and Magic Leap, 3D gestures could really become mainstream interaction modes. Some 3D gesture sets have been designed as part of toolkits and Application Programming Interfaces (APIs) that work with these devices. We are interested in a subset of 3D gestures that focus on tracking hands known as mid-air hand gestures (e.g. without pens and gloves).

Some example mid-air hand gestures for media play back from the Gestoos Software Development Kit[1] are shown in Figure 1. A backwards gesture to move to the previous track or presentation slide is performed by two hands with a *side on* 'T' symbol, while the same symbol performed in the opposite direction is a forwards gesture for moving to the next track or presentation slide. A pause or start gesture is performed with a two hand symbol to play music or videos. A close or quit gesture for the current playing track or ending a presentation is performed with a 'T' symbol. A volume up gesture to increase the sound in a track is performed by placing one hand behind an ear. While a volume down gesture for decreasing the sound is performed by placing one finger over a closed mouth.

A common way to define gesture sets for user interfaces is to conduct user-elicitation studies where participants perform gestures for specific tasks. User defined gestures sets have been developed for surface computing Wobbrock et al. (2009), motion gestures for mobile interaction Ruiz et al. (2011), free-hand TV control Vatavu (2012), multi-display environments Seyed et al. (2012), augmented reality Piumsomboon et al. (2013), and single-hand microgestures Chan et al. (2016). However, very few gesture sets exist for mid-air hand interaction. Nor have many of these mid-air hand gestures been empirically evaluated and reported outside of research labs. There is a need to understand what gestures exist, how to classify gestures, and what gestures work in practice.

---

[1]http://www.gestoos.com/

(a) Backwards

(b) Forwards

(c) Pause/Start

(d) Close/Quit

(e) Volume Up

(f) Volume Down

***Figure 1:*** *Example mid-air hand gestures for media play back from the Gestoos SDK. Semaphoric-Static: Backwards, Forwards, Pause/Start, Close/Quit. Iconic-Static: Volume Up, Volume Down.*

To summarize the research we conducted the first comprehensive *systematic literature review* on 3D mid-air hand gestures following existing systematic literature review methods. Our review aims to present and analyze mid-air hand gesture techniques from the literature. Moreover, the literature review provides a classification of these mid-air gestures and identifies what gestures have been evaluated. The review addresses the following research questions:

**RQ1** What mid-air hand gestures exist in the literature?

**RQ2** What are the gesture classifications for the identified mid-air hand gestures from the literature?

**RQ3** How have the identified mid-air hand gestures from the literature been empirically evaluated?

We first review the existing literature discussing gesture types and classifications, and reviews of gestural interfaces. We then describe the literature review method we used to select and analyze the papers for our review and describe the methodology for analyzing each of the included papers. We present the results and findings of the literature review based on our research questions. Finally, we conclude with some discussions of the results and implications for the design of mid-air hand gestures.

## 2. RELATED WORK

We begin with describing gesture types and classifications, followed by literature reviews conducted on gestural interfaces.

### 2.1. Gesture Types and Classifications

Gestures can be used as an input medium for various types of media ranging from mobiles, to tabletops, and to large wall-displays. The interaction techniques can be direct surface touch, mid-air gestures, full-body gestures or a combination of some or all. Some gestures can be assisted through a physical medium such as a pen or a remote control and others only use parts of the human body such as hands or full-body postures. Gesture types are typically for well-defined actions such as confirmation, selection, navigation and modification. For this study we concentrated on mid-air hand gestures as input without the use of any assisted physical mediums.

Aigner et al. (2012) suggested that many designers tend to focus on a 1:1 relationship mapping between the desired system action and the corresponding mid-air hand gesture to execute the action. This leads to multiple and different mid-air hand gestures for the same action rather than a uniform set of gestures. They have conducted a study, which builds upon a gesture taxonomy model by Karam and Schraefel (2005) in order to create a classification model specific to mid-air hand gestures. They have defined five main classes of gestures and two subclasses which resulted in eight gesture types: Pointing, Semaphoric-Static, Semaphoric-Dynamic, Semaphoric-Stroke, Pantomimic, Iconic-Static, Iconic-Dynamic, and Manipulation (see Table 1).

### 2.2. Reviews and Surveys of Gestural Interfaces

There are different ways in which researchers can reach an understanding of all the literature within a specific domain. A common method is *systematic reviews* which are used to find out what research has been conducted in a specific domain and then to summarise the findings Gough et al. (2013). This type of research highlights the research trends in a specific domain and the possible gaps for new research to be conducted. Surveys are similar to systematic reviews, but are usually more focused on a specific area within a domain. A state of the art study is performed by researchers exploring novel ideas and current trends. The EPPI Centre Methodology[2] is a community concerned with the development of the methods used to conduct systematic reviews in the healthcare domain.

---

[2]http://eppi.ioe.ac.uk/cms/

| Classification Type | Description |
|---|---|
| *Pointing* | Used to isolate a specific object or to specify a specific direction, but does not have to be conveyed by just using the index finger can also be conveyed through multiple fingers or a flat palm. |
| *Semaphoric-Static* | Derive meaning from social symbols or etiquette such as thumbs-up for "OK" or a forward-facing flat palm for "STOP." The symbol can be carried out with one or two hands and is directed at the screen without movement. |
| *Semaphoric-Dynamic* | Similar to semaphoric-static, but are driven by their temporal aspect through continual movement. An example would be the continuous clock-wise motion meaning "ROTATE CLOCKWISE" or the continual flicking of the the hand in a sideways motion indicating "NO." |
| *Semaphoric-Stroke* | Similar to that of semaphoric-dynamic, but differ in that it is a single dedicated stroke. Smartphones and e-readers currently implement this as "NEXT PAGE" or "PREVIOUS PAGE." |
| *Pantomimic* | Think of a mime actor who combines a series of basic actions to illustrate a task. An example is grabbing a document, moving it elsewhere and then dropping it, all performed as a single action. |
| *Iconic-Static* | Pertain to an image or an icon, such as making a circle by cupping two hands together. This differs to semaphoric-static in that it does not have a social or cultural meaning – it can be used to draw a circle on the screen, convey to open a document or anything else that the designers specified the derived meaning should be. |
| *Iconic-Dynamic* | Similar to iconic-static, but requires the movement of both hands to form the outlines of an icon such as a triangle. |
| *Manipulation* | Execute a task as the user performs the gesture. An example is the resizing of an object. There is no delay between the time that the user performs the gesture and the time that the object's size gets updated on the screen. |

***Table 1:*** *Gesture Classification Scheme, defined by Aigner et al. (2012).*

Systematic Literature Reviews (SLR) are a rigorous and methodical review of current literature within a specific domain or subdomain Biolchini et al. (2007); Kitchenham and Charters (2007). SLRs are not only used as a summary of the research area, but also serves as evidence-based guidelines for researchers. Kitchenham et al. (2009) conducted a SLR to determine how popular SLR's are within software engineering research as opposed to using an evidence-based software engineering method. The method for conducting SLR's rests on best practices Kitchenham and Charters (2007).

Some researchers have conducted literature reviews and surveys on gestural interfaces. Erazo and Pino (2015) conducted a literature review and defined a set of operators specifically aimed at the stroke phrase (s-phrase) of mid-air (or touchless) hand gestures. They focused on a specific area of gestures and did not fully discuss the results of their findings from the literature review and did not consider all possibilities for mid-air hand gestures. Shakeri et al. (2014) conducted a SLR on multi-surface (e.g. walls, tabletops, tablets) interactions with geospatial data with the aim to determine which interaction techniques exist in relation to geospatial tasks and activities. They did not consider mid-air hand gestures instead focusing only on devices as input. Liu and Kavakli (2010) conducted a survey on the issues related to multi-model interfaces for computer games and primarily concentrated on speech and hand gestures. Sowa (2008) conducted a survey aimed to provide an overview on advances in hand gesture recognition and comprehension as used in HCI, especially focusing on how users relate to coverable (accompanying speech) gestures in human communication. Wu and Huang (1999) conducted a review of early vision-based gesture recognition techniques and systems, which did not include the ubiquitous devices we see today.

Our analysis identified that there is no recent comprehensive review of existing mid-air hand gestures. To gain a greater understanding of what mid-air gestures with interactive surfaces and displays exist in the literature and to address our research questions (RQ1–3) we conducted a SLR.

## 3. LITERATURE REVIEW: METHOD

We conducted a systematic literature review following existing methods as they have been successfully applied in HCI and Software Engineering Biolchini et al. (2007); Gough et al. (2013); Kitchenham et al. (2009); Kitchenham and Charters (2007). The steps of the review method are now outlined.

### 3.1. Step 1. Research Process

We performed a search for relevant papers using Google Scholar and keywords, see Table 2. We excluded patents. The keywords included combinations of the following terms: mid-air hand gesture, mid-air gesture, 3D gesture, in-air gesture, freehand/barehand gesture, interaction, pen, touch, touchless, MS Kinect, Leap Motion, Asus Xtion, Creative Senz3D, vicon, and motion capture. Phase 1 of the search used keywords and 35 papers were found. Phase 2 used backward snowballing from the references listed in the papers in Phase 1 and 18 papers were found. Phase 3 used forward snowballing to find citations of the papers from Phase 1 and 12 papers were found. Giustini and Boulos (2013) found that Google Scholar alone was not enough to comprehensively conduct a SLR. For this reason we searched electronic databases (e.g. ACM Digital Library, IEEE Xplore, Springer, ScienceDirect) and HCI conferences / journals (e.g. CHI, ITS, UIST, AVI, 3DUI, VR, UbiComp, ISMAR, Interact, MobileHCI, and regional HCI/UI conferences) to find relevant papers.

## 3.2. Step 2. Inclusion and Exclusion Criteria

A paper was included for evaluation based on one or more of the following criteria: described or proposed a set of mid-air hand gestures, evaluated a mid-air hand gesture set, evaluated the effectiveness of mid-air hand gestures for specific tasks, or compared a set of mid-air hand gestures. Any paper that mentioned mid-air hand gestures, but did not describe or evaluate the gestures were excluded. As our focus was on mid-air hand gestures only we excluded papers which evaluated mid-air gestures with the aid of devices (e.g. pen or wand) or papers that only concentrated on full body gestures (e.g feet, legs, and torso). Papers that contained mid-air hand gestures as part of the full body gesture set, were included, but only the mid-air hand gestures parts.

## 3.3. Step 3. Quality Assessment

We each individually analysed and classified the set of papers. The results were updated in an online collaborative spreadsheet. All excluded papers were clearly marked and detailed reasons were given on why a paper was excluded. Regular meetings were held between all the researchers where the mid-air hand gestures were discussed and classifications examined. When there was a disagreement about a classification, we discussed the matter until a consensus was reached on which classification to use for a specific gesture.

## 3.4. Step 4. Data Collection and Analysis

We extracted the initial list of papers and each paper was evaluated as follows.

*Identification and Overview:* We used the title, keywords, publication year, authors, conference title and the paper's source to identify and differentiate between each. As we evaluated each paper we wrote a short summary to provide an overview of the research conducted.

*Inclusion or Exclusion:* For each paper we captured the devices they used to perform mid-air hand gestures (if any). This information together with the summary information gave us enough data to decide if the paper matched our criteria for inclusion or exclusion. If the paper was excluded, we wrote an explanation on why the paper was excluded for use within the discussion phase.

*Research Method:* We were interested in knowing what type of research evaluation method was used in carrying out each study, to give an indication of how rigourosly the mid-air hand gestures were evaluated. We also captured the input precision of each gesture to differentiate between gestures for low precision

| Phase | Method | # Papers |
|---|---|---|
| 1 | Search Keywords | 35 |
| 2 | Backward Snowballing of Phase 1 | 18 |
| 3 | Forward Snowballing of Phase 2 | 12 |
| | **Total papers** | 65 |

**Table 2:** *Research Process: phase of search iteration, methods applied, and number of papers included.*

(e.g. manipulation of 3D blocks) or high precision (e.g. manipulation of text).

*Snowballing:* At the end of each phase, we discussed the results and then examined the references (backwards snowballing) of each paper to determine if we could include more papers which matched our inclusion and exclusion criteria. We subsequently examined the citations of these papers (forward snowballing) to see if any papers could be included.

*Gestures:* For each paper we included, we captured the specific task that the gesture aimed to perform (e.g. open menu). We categorized each task according to how many hands were involved in performing the gesture (e.g. 1 or 2 hands). Additionally we classified each mid-air gesture task according to the classification scheme by Aigner et al. (2012), see Table 1. The first level of the scheme is divided into five major classes based on the usage of the gesture while a second level shows physical appearances and properties of each gesture (e.g. static and dynamic). The gesture scheme was adapted for mid-air hand gestures based on work by Karam and Schraefel (2005).

## 3.5. Step 5. Data Coding and Synthesis

To extract codes from the papers we applied a version of the grounded theory methodology to extract concepts based on Strauss and Corbin (1997). This process involved reading the papers to extract concepts. We coded each data point of the text in the paper that was relevant to the scope of the review. We next read all of the highlighted words, sentences and expressions and categorized them base on their similarities. This process of generating higher-abstraction level type categories from a set of concepts is called open coding Wolfswinkel et al. (2013). We then created axial codes to represent the relationship and links between categories and their subcategories. For each paper the analytical coding steps were performed iteratively until theoretical saturation was reached which is when no new information, concept, or relationship emerges from the data. We now present the results from conducting the different steps of the SLR. Table 3 shows the results of the open coding and axial coding steps.

## 4. LITERATURE REVIEW: RESULTS

In this systematic literature review we identified 65 papers and analysed the papers in detail to address the research questions RQ1–RQ3. The papers in the review are listed in the Systematic Literature References section (§8).

### 4.1. Overview of Papers

Figure 2 shows the area of venue of the papers and publication year. The largest conference venue was CHI which had eight papers (P2, P3, P12, P16, P18, P34, P53, P64). We found 47 papers in other HCI related and regional conferences (e.g. UIST, CHI-EA, ITS, MobileHCI, Interact, AVI, PerDis, NordiCHI, HCI International, UbiComp, Multimedia, OzCHI, GI, CHItaly). We found some papers in other domains: Engineering (4), Graphics (3), Science (2), and Medicine (1). Prior to 2011 there were five papers that focused on mid-air hand gestures. From 2011 and onwards there has been a steady increase of papers. This indicates that 3D sensor devices were more accessible within research groups and probably likely due to low cost devices being available such as the MS Kinect (2010). 2013, 2014, and 2015 shows 16, 19, and 7 papers respectively, then only one paper to date has be found for 2016.
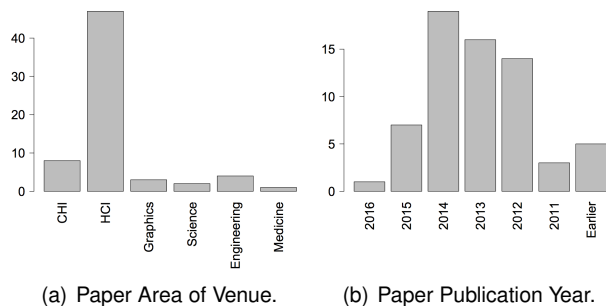


(a) Paper Area of Venue.    (b) Paper Publication Year.

**Figure 2:** *Overview of literature review papers.*

### 4.2. RQ 1 - Mid-Air Hand Gestures

In order to answer RQ1 of this review, we use the output of the grounded theory process to present a summary of results in Table 3. This table represents three levels of the grounded theory approach: concepts, open codes, and axial codes (categories of open codes). In this section we discuss each of these categories combined with their related codes and concepts in more detail. We discuss the categories in most to least frequent in terms of number of papers per category.

#### 4.2.1. Selection
This category includes concepts related to selecting items, performing executions, selecting control options for media, and confirming selections.

*Item:* This code refers to how items (including menu items) can be selected. 47 papers focused on item selection gestures. We grouped all gestures together which consisted of highlighting, selecting, and activating an item, as these tasks were considered to be related. One paper focused on mid-air text entry gestures which was grouped under the Highlight concept, as the activation of an item revealed further selectable hot-spots. This was considered to be a different highlight action. Gestures in a different paper focused on the selection of chunks of items and their sub-items by tilting the wrist in a clockwise or anti-clockwise direction. All items were activated and highlighted through Pointing gestures, Semaphoric Static/Stroke/Dynamic gestures and Iconic Static gestures. The Cancellation of a selection was achieved predominantly by using Semaphoric Stroke actions.

*Command Execution:* This code refers to commands that execute a particular action within an application. 15 papers focused on Start and Open commands which were for starting up a system or opening an application. These commands were performed with Pointing, Iconic Static/Dynamic, Semaphoric Static/Dynamic/Stroke gestures. The associated commands included End/Close and Restart/Reset events, which predominantly used Pointing and Semaphoric Stroke gestures. Two papers covered Identification and Authentication commands which consisted predominantly out of Iconic Dynamic gestures as it required the user to draw a letter or a unique pattern in the air. The rest of the commands covered actions such as calling up a Help Menu and Show, Hide, Lock and Unlock actions which consisted of Semaphoric Stroke/Dynamic gestures. One medical paper was able to set measurement callipers using two hands and was performed using a Semaphoric Dynamic gesture.

*Media Control:* This code refers to actions mainly used to control media for video, audio, and phone. 14 papers focused on gestures for Play, Pause, Skip, Start, Stop, Fast Forward, Go Backward and Record actions. These actions consisted of Semaphoric Stoke and Semaphoric Dynamic gestures. One paper mentioned a gesture for "breaking into" a live message recording which meant that a voice message being left by the sender, could be interrupted by the recipient.

*Confirmation:* This code refers to Accept (e.g. yes and agree) and Refuse (e.g. no) actions and included Pointing and Semaphoric Static/Dynamic/Stroke gestures. 12 papers focused on Accept gestures and five papers covered both Accept and Refuse gestures.

| Cat. | Open Code | Concepts |
|---|---|---|
| **Selection (179)** (# of papers: 63) | **Item (47)** (P1, P4, P5, P6, P7, P8, P11, P12, P14, P15, P17, P19, P20, P24, P26, P27, P28, P30, P31, P32, P36, P37, P38, P39, P42, P43, P44, P45, P49, P51, P52, P53, P54, P57, P58, P59, P63, P65) (# of papers: 38) | **Activation (38)** [P1, P4, P5, P6, P7, P8, P11, P12, P14, P15, P17, P19, P20, P24, P26, P27, P28, P30, P31, P32, P36, P37, P38, P39, P42, P43, P44, P45, P49, P51, P52, P53, P54, P57, P58, P59, P63, P65], **Global/Context/Sub Menu Item (4)** [P5, P6, P54, P58], **Group of Items (3)** [P11, P42, P53], **Highlight (1)** [P37], **Arrange (1)** [P52] |
| | **Command Execution (61)** (P6, P7, P10, P15, P17, P20, P22, P26, P32, P33, P34, P35, P37, P39, P42, P44, P45, P49, P50, P51, P52, P53, P54, P56, P57, P58, P62, P64) (# of papers: 28) | **Start/Open/On/New (15)** [P6, P10, P20, P32, P35, P39, P42, P45, P49, P51, P53, P57, P58, P62, P64], **End/Close/Off (12)** [P10, P17, P35, P42, P44, P45, P50, P51, P52, P53, P56, P58], **Cancel/Undo/Redo/Escape (6)** [P7, P26, P42, P49, P53, P54], **Restart/Reset/Return (5)** [P33, P35, P37, P56, P57], **Help (4)** [P10, P42, P53, P58], **Insert/Cut/Copy/Paste/Delete/Duplicate (3)** [P42, P44, P53], **Show Action (2)** [P7, P42], **Identification/Authentication (2)** [P22, P34], **Find (2)** [P42, P53], **Save (2)** [P42, P53], **Minimise/Maximise (2)** [P42, P53], **Hide Action (2)** [P7, P58], **Lock Selection (1)** [P7], **Unlock Screen (1)** [P15], **Set Measurement Callipers (1)** [P17], **Switch Task (1)** [P42] |
| | **Media Control: Video, Audio, Phone (54)** (P16, P18, P21, P41, P42, P44, P45, P50, P51, P53, P55, P58, P61, P64) (# of papers: 14) | **Go Backward/Previous/Next/Favourite/Random/Last Visited (9)** [P18, P41, P42, P45, P51, P53, P58, P61, 64], **Play/Resume (7)** [P18, P41, P50, P51, P53, P55, P61], **Increase/Decrease/(Un)Mute Sound/Tempo (7)** [P21, P42, P45, P51, P53, P58, P61, P64], **Fast Forward (5)** [P18, P41, P42, P45, P51], **Pause (5)** [P18, P50, P51, P53, P61], **Stop (4)** [P18, P41, P53, P61], **Skip Scene (2)** [P18, P53], **Record video message (1)** [P16], **Stop recording a video message (1)** [P16], ( **Break into live video chat (1)** [P16], **Check Messages (1)** [P44], **Answer Call (1)** [P44], **Ignore Call (1)** [P44], **Place on Hold (1)** [P44], **End Call (1)** [P44], **Place Call (1)** [P44], **Check Calendar (1)** [P44] **Protest Song (1)** [P64] **Equip with (1)** [P64] **Aim (1)** [P64] **Throw (1)** [P64] **Change equipment (1)** [P64] |
| | **Confirmation (17)** (P1, P5, P8, P10, P13, P33, P35, P42, P44, P53, P55, P58) (# of papers: 12) | **Accept (12)** [P1, P5, P8, P10, P13, P33, P35, P42, P44, P53, P55, P58], **Refuse (5)** [P10, P42, P44, P53, P58] |
| **Navigation (88)** (# of papers: 40) | **Zoom (29)** (P2, P3, P4, P15, P17, P20, P29, P32, P40, P42, P43, P46, P47, P48, P49, P50, P52, P53, P54) (# of papers: 18) | **Zoom In (Linear/Circular) (18)** [P2, P3, P4, P15, P17, P20, P29, P32, P40, P42, P43, P47, P48, P49, P50, P52, P53, P54], **Zoom Out (Linear/Circular) (11)** [P2, P4, P15, P42, P43, P47, P48, P49, P50, P53, P54] |
| | **Scroll (25)** (P1, P4, P7, P10, P15, P16, P17, P20, P21, P26, P29, P32, P33, P35, P43, P44, P49, P53, P55, P57, P63) (# of papers: 21) | **Move left/right (17)** [P1, P4, P10, P16, P17, P20, P21, P26, P29, P32, P35, P43, P44, P49, P53, P55, P57], **Move up/down (8)** [P7, P15, P20, P33, P35, P43, P49, P63] |
| | **Move Cursor (15)** (P3, P6, P7, P12, P17, P20, P23, P26, P31, P32, P37, P39, P40, P42, P56) (# of papers: 15) | **Move cursor up, down, left, right (14)** [P6, P7, P12, P17, P20, P23, P26, P31, P32, P37, P39, P40, P42, P56], **Handle Bar - Translate, Rotate, Extend, Modify (1)** [P3] |
| | **Pan (10)** (P3, P15, P18, P42, P43, P47, P48, P49, P50, P53, P54) (# of papers: 11) | **Point and Move (6)** [P42, P47, P48, P50, P53, P54], **Grab and Move (3)** [P15, P43, P49], **Joystick4D/Joystick360 (1)** [P47] |
| | **View Control (9)** (P3, P18, P19, P31, P40) (# of papers: 5) | **View Rotate (2)** [P3, P40], **Pan (2)** [P3, P18], **Zoom (1)** [P18], **Move Left/Right (1)** [P19], **Move Forward/Backwards (1)** [P19], **Look Up/Down (1)** [P19], **Toggle Camera Position ((1)** [P31] |
| **Manipulation (39)** (# of papers: 18) | **Single Object (SO) (32)** (P3, P9, P20, P25, P26, P27, P28, P31, P33, P42, P46, P48, P49, P50, P52, P53, P56, P60) (# of papers: 18) | **Rotate (12)** [P3, P20, P31, P33, P42, P46, P48, P49, P50, P53, P56, P60], **Translate (8)** [P20, P25, P26, P31, P42, P52, P53, P56], **Scale (6)** [P3, P9, P31, P42, P52, P56], **Change State (3)** [P27, P28, P60], **Extrude (1)** [P9], **Set Extrude Height Constraint (1)** [P9], **Constrained Rotation (1)** [P3] |
| | **Multiple Objects (7)** (P3, P9) (# of papers: 2) | **Scale (2)** [P3, P9], **Select Active Objects (1)** [P3], **Rotate (1)** [P3], **Alignment (1)** [P3], **Extrude (1)** [P9], **Set Extrude Height Constraint (1)** [P9] |

**Table 3:** *Mid-Air Hand Gestures Overview – categories, open codes, and concepts extracted from the included papers in this literature review. Numbers in bold parentheses show the number of occurrences of each concept and code from the included papers.*

### 4.2.2. Navigation

This category includes concepts related to navigating with an interface such as zooming, scrolling, panning, and moving a cursor around the screen.

*Zoom:* This code refers to zooming in and out of the view and can be performed by Linear or Circular actions. Linear refers to straight movements such as pushing a hand towards the screen to zoom in. Circular zooming required the user to move their hand in a clockwise direction to zoom in. 18 papers focused on gestures to Zoom in either Linear or Circular and 11 papers covered Zoom Out Linear or Circular gestures. Linear gestures were covered with Semaphoric Stroke / Dynamic gestures whereas circular gestures covered only Semaphoric Dynamic gestures. Three papers used two hands to perform the zooming, where one hand was used to point to the active object and the other controlled the zooming action.

*Scroll:* This code refers to the scrolling action which allows the user to move between items on the screen (such as photographs or menu items). We found that scrolling was covered on both the horizontal and vertical axes. 17 papers focused on left/right scrolling and eight papers up/down scrolling. Semaphoric Stroke and Semaphoric Dynamic gestures dominated the scrolling action.

*Move Cursor:* This code refers to moving some form of a cursor on screen. 14 papers created gestures to move the cursor up, down, left and right on the screen. One paper used a cursor in the form of a handlebar, which represents a bar that can be moved around the screen in much the same way a generic cursor can. The default orientation of the Handle Bar is horizontal, but it can be rotated in any orientation. The Handle Bar can be extended to become a much larger cursor. The Handle Bar cursor was mainly used to indicate the pivot point and orientation which influenced other actions such as the axis by which to perform a rotation.

*Pan:* This code refers to moving the view port to a different part of the screen. Six papers described the panning action of Point and Move, and three papers for Grab and Move. These gestures were performed with Manipulation gestures.

*View Control:* This code refers to controlling the camera view within a virtual environment or an application. Two papers focused on rotating the view. Two papers used the camera view to zoom in or to pan. One paper focused on moving the camera left, right, up, down, backwards, or forwards. One paper allowed the user toggle the camera's position between full and 3/4 views. There gestures were

performed with either Manipulation or Semaphoric Dynamic gestures.

### 4.2.3. Manipulation

This category includes concepts related to the manipulation of 2D and 3D objects. We have divided this group into two sections, namely Single Objects and Multiple Objects, as the process of manipulating multiple objects differs to that of Single Objects. The actions are, however, very similar.

*Single Object:* 16 papers focused on performing rotation, translating, and scaling operations on an object. Other activities included changing the state of an object and extruding an object. One of the papers allowed the user to lock in the rotation's pivot point by making use of a configurable Handle Bar cursor. A different paper allowed the user to point with one hand to where the extrusion height should be limited to. In a virtual environment, one paper allowed the user to change the state of an object such as moving the hand forward to open a window and pulling the hand away from the window object, to close it again. These gestures were performed with Manipulation and Semaphoric Dynamic gestures although a few used Pantomimic gestures which used drag and drop or enclose and drop gesture sets.

*Multiple Objects:* Two papers focused on manipulating multiple objects at once. Manipulation with multiple objects were similar to that of Single objects, except that all gestures required two hands to perform the task. The scaling gesture was most prevalent followed by selection, rotation, extrude, and alignment gestures. The biggest difference between multiple object and single object manipulation was the ability to align objects and to select which objects were active. All actions consisted of Manipulation gestures and one Semaphoric Static gesture which was used to point to where the extrusion height limit was.

### 4.3. RQ 2 - Mid-Air Hand Gestures Classification

Table 1 describes a gesture classification scheme by Aigner et al. (2012) which we used to classify the gestures from the papers. Figure 3 shows the results of classifying 485 gestures. The gesture classifications are labelled along the X axis and frequency along the Y axis. The majority of gestures were from the selection category (327 gestures) followed by navigation (109 gestures) and then manipulation (49 gestures). Gestures for selection were mainly Semaphoric-Stroke and Semaphoric-Dynamic actions; gestures for navigation were mainly Semaphoric-Dynamic and Semaphoric-Stoke actions; and finally gestures for manipulation were performed with Manipulation actions.

*Pointing:* This classification was predominantly used for Selection tasks followed by Navigation tasks. We did not find any Manipulation tasks under the Pointing Classification. This may be due to the nature of Manipulation which changes the physical aspects of the object and is thus not directly related to Pointing.

*Semaphoric Static:* This classification is similar to pointing and covered mostly the Selection tasks and two Manipulation tasks. These two Manipulation tasks used the non-dominating hand to indicate where the limit for object extrusion point should be. We did not place these gestures under Pointing as showing the limit has cultural meaning and therefore we classified them under Semaphoric Static.

*Semaphoric Dynamic:* Gestures within this classification were mostly associated with Selection tasks, but was also used for Navigation tasks. Moving the hand in both directions was used for selecting items, turning devices on and off, handling media controls and confirmation tasks. Some Manipulation tasks included the wave gesture to rotate objects.

*Semaphoric Stroke:* The Selection category was most common within this classification especially for activating menu items and for performing media control tasks such as selecting next and previous items on the screen. Navigation also used semaphoric-stroke gestures for scrolling, zooming, and view control tasks. Manipulation tasks which changed the object's state also fell under this classification as a door was pushed open or a window was pulled shut.

*Pantomimic:* Not many tasks were associated with the Pantomimic classification. Gestures within the Manipulation category made use of it by grabbing an object, placing it elsewhere and then letting go of it. For zooming in and out under Navigation, the user would grab an item and then pull it away from the screen or grab it and push it toward the screen. Under the Selection category, the user would use two fingers in a scissor-like gesture followed with a downward push for insertion and duplication tasks.

*Iconic Static:* We found only two Iconic Static gestures within the Selection category, where an "O" shape was used to invoke speech recognition and an "L" shape to select an item.

*Iconic Dynamic:* Similar to Iconic Static, this gesture classification was primarily used within the Selection category. Most of these gestures were for Identification and Authentication tasks where the user was required to draw multiple symbols, numbers, and letters. A few gestures were also related to gestures above mobile phones. The user
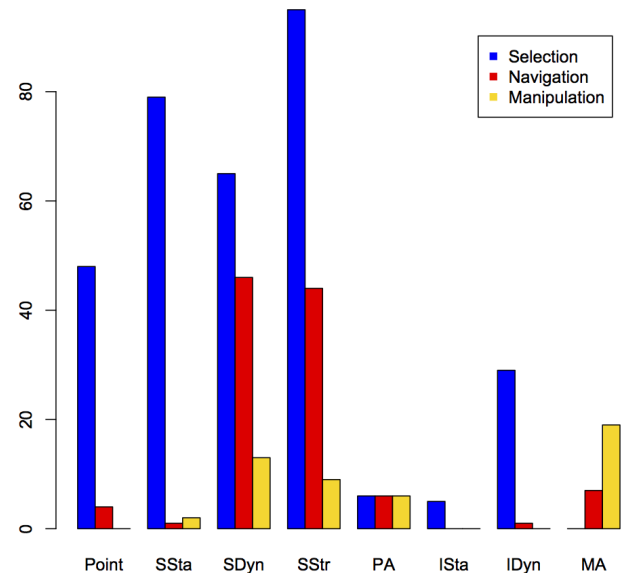


**Figure 3:** *RQ2 - Gesture Classification Results from the papers identified in the literature review. Gestures in order from left to right: Pointing, Semaphoric-Static, Semaphoric-Dynamic, Semaphoric-Stroke, Pantomimic, Iconic-Static, Iconic-Dynamic, and Manipulation.*

drew a square to read messages, a question mark to check the calendar and an X to delete items.

*Manipulation:* The Manipulation category comprised mostly of Manipulation classifications where objects were rotated, translated, scaled, and extruded. For the Navigation category, most of the gestures were used to manipulate the View Control on the X, Y, and Z axes.

We found that most gestures were performed using only one hand (uni-manual) and that less than 20% of the gestures required two hands (bi-manual) to perform an action. Two handed gestures (bi-manual) were mostly used for Semaphoric-Dynamic, Manipulation, and Semaphoric-Stroke gestures. One paper used two hands to perform Pantomimic actions with their Handle Bar concept. A different paper made it possible to select menu items by using either one or two hands and by showing the corresponding finger count that corresponds to the number on a menu item.

### 4.4. RQ 3 - Mid-Air Hand Gestures Evaluation

We wanted to know what methods were performed to evaluate the mid-air hand gestures, how many people participated in the studies, and what kind of camera detection devices were used.

The papers used a variety of study methods to perform evaluations of the gestures. These study types include elicitation (similar to Wobbrock et al. (2009)), Wizard of Oz (similar to Kelley (1984)),
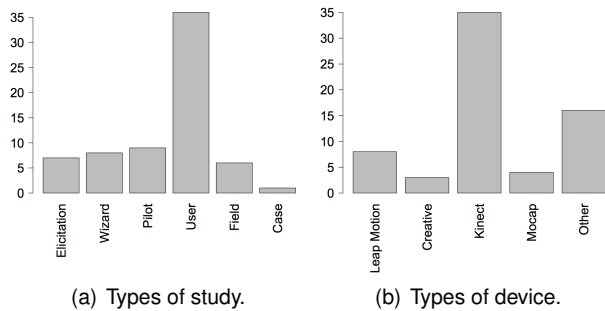
(a) Types of study.

(b) Types of device.

**Figure 4:** *RQ3 - Types of evaluation performed and types of devices used for gesture detection.*

pilot study (performed during design), user study (performed in the lab), and field study (performed in the wild). Figure 4(a) shows the different study methods used. Most papers conducted a user study (36 papers). The number of participants in the user studies ranged from 5–70 and total of 740 participants. Nine papers performed pilot studies, and had a total of 94 participants. Eight papers performed Wizard of Oz studies and had a total of 144 participants. Six papers performed field studies with a total of 1553 participants. Two of these field study papers had a significant number of participants P8 (455) and P39 (approx. 1000). Seven papers performed elicitation studies and had a total of 116 participants. The combined total number of studies conducted was 72 with a combined total of 2647 participants. One paper performed a case study illustrating their techniques on a real application. Five papers did not have any evaluations performed at all. A number of papers performed multiple studies and different types of studies. With regard to different types of studies some papers (P1, P6) performed a Wizard of Oz study followed by a user study or field study, two papers (P27, P35) conducted an elicitation study followed by a user study, and one paper (P34) conducted a user study followed by a field study.

Each of the papers used different devices for tracking and detecting gestures. Figure 4(b) shows the different devices starting from short range to longer range along the X axis. Overwhelming most papers (35 papers) used the MS Kinect[3]. Eight papers (P4, P23, P46, P48, P54, P58, P62, P65) used the Leap Motion[4] device and five studies were conducted in 2014 and three in 2015. Three papers (P12, P22, P28) used the Creative Senz3D[5] camera, and these studies were conducted in either 2013 or 2014. Four papers (P2, P7, P11, P37) used motion capturing

---

[3]https://www.microsoft.com/en-us/kinectforwindows/
[4]https://www.leapmotion.com/
[5]http://uk.creative.com/p/web-cameras/creative-senz3d

systems such as Vicon[6] cameras, and these studies were conducted in 2011 (x2), 2013, and 2014. 16 papers used other devices such as undefined depth sensing cameras, Firefly Point Grey camera, wrist bands, or custom made tracking cameras. One paper (P9) used a MS Kinect in combination with a Game Track device to detect additional features of the hands of participants.

## 5. DISCUSSION

We conduced a large comprehensive SLR of mid-air hand gestures with interactive surfaces and displays. Related reviews addressed a smaller number of papers, was prior to 2011 (and MS Kinect), or did not present any analysis of their findings.

Most mid-air hand gestures were researched for Item Activation, Zoom and Scroll, Move Cursor and Accept concepts. This indicates that research is aimed towards the most common interaction techniques, perhaps with the aim to augment traditional mouse and keyboard interaction as sensor technology becomes more widely available. It was interesting to see that contrasting concepts such as Accept and Reject, Zoom In and Out, and Scroll Horizontal and Vertical did not yield equal frequencies. This may be due to how research tasks are being constructed for user studies. Not much research has been performed on text manipulation, which may indicate that sensors for mid-air hand gesture interaction are not precise enough yet to cater for this task.

We found two studies which investigated text input. One study (P66) required participants to free-type. Free-type is when the user is able to type on a keyboard without having to look at where the keys are located. The Leap Motion sensors in this study were able to detect finger movements in the air and how they corresponded to the key placements on a keyboard. The second study (P37) that was included, used concepts of mobile phones such as the reduced 9-key keyboard, H4-Mid-Air and QWERTY. P37 had most success with the QWERTY layout at 13.2 WPM (Words per Minute), although input was with only one hand. P65 had a bit more than double the rate at 29.2 WPM, but they made use of both hands and the users were experienced in free-type. Both studies admitted that more research was required in this area which suggests that users are not quite ready to give up their physical keyboards. Since not all users are able to free-type, researchers might want to consider other types of text input such as what P37 explored and may even want to adapt touch screen text entry techniques such as word swipes with predictive wording.

---

[6]http://www.vicon.com/

Studies conducted took place in a few distinctive domains including (but not limited to), medical, public display interaction, media content control (drawing, video, music and gaming) and smart object control. The requirements for each domain were significantly different. For example, the medical domain required more finely tuned manipulation of medical displays than what would be expected from public display or media control interactions. There were also significantly less pantomimic gestures which suggest that users are very much embedded in controlling items by means of button pushing such as found with WIMP based interfaces. The gaming industry might be more inclined to try these gestures as it may enhance the interactive and enjoyment experience, but even they need to consider the physical constraints it poses on the human body. For example, a user could easily spend an hour emerged in gameplay using a mouse and keyboard whereas using pantomimic interactions (e.g. throwing item) might cause fatigue.

Gesture classifications showed that Semaphoric Stroke/Dynamic and Pointing classifications are the most widely researched classifications within the Selection and Navigation categories. This would indicate that simplistic gestures are preferred over complex combinations. This may be due to users familiarity of WIMP based interfaces. The least researched classifications were Iconic Static and Pantomimic classifications. This may indicate that gestures within Iconic Static and Pantomimic classifications were too difficult or mentally taxing for users to consider and remember. Most work in this area has been done in the media content control domain where menu selection played a large role. Counting appeared to be popular using different techniques of hierarchy and segmentation and although this could be considered appropriate for large displays in public areas, it seems rather unlikely that this method would be adopted as practical in a business environment, where it could be much quicker to use traditional input methods. Many interactions were also related to executing specific commands such as opening and closing of content. These are similar actions seen on menu bars in an application, suggesting a gradual intention to move away from the traditional input methods rather than a complete redesign of the current look and feel of applications. Several studies were aware of the practicality of gestures in regards to the physical space available to make the gesture and possibly the embarrassment that may accompany it when performing the gesture. These studies concentrated on subtle hand movements where finger pinching and sliding proved successful. Having said that, these studies did require special sensors to pick up the subtle muscle movements of the hand. These

types of research may prove to be more practical for the business environment where discretion is expected or where the user is unable to make elaborate gestures. It was interesting to see that one study attempted to use mid-air gestures for authentication control (P22) and that they were rather successful with it.

Although most smart object papers were excluded, such as ones that used motion capturing systems like Vicon cameras, which suggests that users are starting to move away from the concept of controlling everyday objects (such as lights in a room) via intermediary equipment (such as switches and remote controls). If more everyday object interaction is via mid-air gestures, then it could make the concept of controlling displays with these gestures seem more usable without the use of keyboards and mice.

Evaluations of mid-air hand gestures showed that the majority of the papers performed a user study from 2011 onwards and used a MS Kinect. Four papers used large expensive setups such as motion capture to detect gestures. Half of the "other" category used devices that existed before the Kinect. While some of the newer papers from the "other" category used devices such as wrist bands or cameras built into TVs, and the Leap Motion.

## 6. CONCLUSION

Understanding what mid-air hand gestures exist can help designers build better user experiences for interactive surfaces and displays. We conducted a comprehensive systematic literature review on mid-air hand gestures and identified 65 papers. To answer our research questions, we analyzed all of the papers using open coding to extract categories and concepts related to mid-air hand gestures, and classified each of the gestures. Our findings showed that mid-air hand gestures have primarily been designed for selection, navigation, and manipulation tasks. For selection the main types of gesture classifications were performed using Semaphoric-Stroke, Semaphoric-Dynamic, and Pointing actions, while for Navigation it was Semaphoric-Dynamic and Semaphoric-Stoke actions, finally manipulation was performed with Manipulation actions. We identified the types of evaluations that were performed on these gesture sets and found that surprisingly only a small amount of gestures have been evaluated and most within a research lab setting. In order for mid-air hand gestures to become ubiquitous more evaluations are required to understand what gestures are effective for what tasks.

## 7. ACKNOWLEDGEMENTS

## 8. SYSTEMATIC LITERATURE REFERENCES

P1. R. Walter, G. Bailly, N. Valkanova, and J Müller. Cuenesics: Using mid-air gestures to select items on interactive public displays. In *Proc. of MobileHCI*, pp. 299–308. ACM, 2014.

P2. M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay. Mid-air pan-and-zoom on wall-sized displays. In *Proc. of CHI*, pp. 177–186. ACM, 2011.

P3. 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 CHI*, pp. 1297–1306. ACM, 2012.

P4. N. Rossol, I. Cheng, R. Shen, and A. Basu. Touchfree medical interfaces. In *Proceedings of the International Conference on Engineering in Medicine and Biology Society (EMBC)*, pp. 6597–6600. IEEE, 2014.

P5. S. Lenman, L. Bretzner, and B. Thuresson. Using marking menus to develop command sets for computer vision based hand gesture interfaces. In *Proc. of NordiCHI*, pp. 239–242. ACM, 2002.

P6. G. Bailly, R. Walter, J. Müller, T. Ning, and E. Lecolinet. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *Proc. of Interact*, pp. 248–262. Springer-Verlag, 2011.

P7. D. Vogel and R. Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. of UIST*, pp. 137–146. ACM, 2004.

P8. M. Perry, S. Beckett, K. O'Hara, and S. Subramanian. Wavewindow: Public, performative gestural interaction. In *Proc. of ITS*, pp. 109–112. ACM, 2010.

P9. D. De Araùjo, G. Casiez, and J. Jorge. Mockup builder: Direct 3D modeling on and above the surface in a continuous interaction space. In *Proc. of GI*, pp. 173–180. Canadian Information Processing Society, 2012.

P10. R-D. Vatavu. User-defined gestures for free-hand tv control. In *Proc. of European Conference on Interactive Tv and Video (EuroiTV)*, pp. 45–48. ACM, 2012.

P11. T. Ni, D. Bowman, C. North, and R. McMahan. Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *Int. J. Hum.-Comput. Stud.*, 69(9): pp. 551–562, August 2011.

P12. A. Kulshreshth and J. LaViola, Jr. Exploring the usefulness of finger-based 3D gesture menu selection. In *Proc. of CHI*, pp. 1093–1102. ACM, 2014.

P13. M. Zarzuela, F. Pernas, S. Calzn, D. Ortega, and M. Rodrguez. Educational tourism through a virtual reality platform. *Procedia Computer Science*, 25(0):382–388, 2013. In *Proc. of Conference on Virtual and Augmented Reality in Education*.

P14. F. van de Camp, A. Schick, and R. Stiefelhagen. How to click in mid-air. In Norbert Streitz and Constantine Stephanidis, editors, *Distributed, Ambient, and Pervasive Interactions*, volume 8028 of *Lecture Notes in Computer Science*, pp. 78–86. Springer, 2013.

P15. A. Tuntakurn, S.S. Thongvigitmanee, V. Sa-ing, S. Hasegawa, and S.S. Makhanov. Natural interactive 3D medical image viewer based on finger and arm gestures. In *Proc. of Biomedical Engineering International Conference (BMEiCON)*, pp. 1–5, 2013.

P16. J. Tang, R. Xiao, A. Hoff, G. Venolia, P. Therien, and A. Roseway. Homeproxy: Exploring a physical proxy for video communication in the home. In *Proc. of CHI*, pp. 1339–1342. ACM, 2013.

P17. J. Tan, C. Cherng, M. Zawaideh, A. Roberts, and T. Kinney. Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *RadioGraphics*, 33(2):E61–E70, 2013.

P18. G. Rovelo Ruiz, D. Vanacken, K. Luyten, D. Abad, and E. Camahort. Multi-viewer gesture-based interaction for omni-directional video. In *Proc. of CHI*, pp. 4077–4086. ACM, 2014.

P19. G. Re and M. Bordegoni. A natural user interface for navigating in organized 3D virtual contents. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, volume 8525 of *Lecture Notes in Computer Science*, pp. 93–104. Springer, 2014.

P20. A. Viveros and E. Hernández Rubio. Kinect, as interaction device with a tiled display. In *Proc. of HCI International*, pp. 301–311. Springer, 2013.

P21. A. Arif, W. Stuerzlinger, E. Filho, and A. Gordynski. Error behaviours in an unreliable in-air gesture recognizer. In *Proc. of CHI EA*, pp. 1603–1608. ACM, 2014.

P22. M. Aumi and S. Kratz. Airauth: Towards attack-resilient biometric authentication using in-air gestures. In *Proc. of CHI EA*, pp. 1585–1590. ACM, 2014.

P23. M. Brown, W. Stuerzlinger, and E. Filho. The performance of un-instrumented in-air pointing. In *Proc. of GI*, pp. 59–66. Canadian Information Processing Society, 2014.

P24. D. Chattopadhyay and D. Bolchini. Touchless circular menus: Toward an intuitive ui for touchless interactions with large displays. In *Proc. of AVI*, pp. 33–40. ACM, 2014.

P25. L. Cohen, S. Haliyo, M. Chetouani, and S. Regnier. Intention prediction approach to interact naturally with the microworld. In *Proc. of Advanced Intelligent Mechatronics (AIM)*, pp. 396–401. IEEE, July 2014.

P26. L. Hespanhol, M. Tomitsch, K. Grace, A. Collins, and J. Kay. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proc. of PerDis*, pp. 6:1–6:6. ACM, 2012.

P27. K. Ho and H. Weng. Favoured attributes of in-air gestures in the home environment. In *Proc. of OzCHI*, pp. 171–174. ACM, 2013.

P28. M. Hosseini, P. Vallotton, T. Bednarz, and A. Sowmya. A study of touchless versus touch-based interactions with bacterial biofilm images. In *Proc. of Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI)*, pp. 285–290. ACM, 2013.

P29. S. Jalaliniya, J. Smith, M. Sousa, L. Büthe, and T. Pederson. Touch-less interaction with medical images using hand & foot gestures. In *Proc. of UbiComp*, pp. 1265–1274. ACM, 2013.

P30. B. Bossavit, A. Marzo, O. Ardaiz, and A. Pina. Hierarchical menu selection with a body-centered remote interface. *Interacting with Computers*, 2013.

P31. Y. Kim, S. Sim, S. Cho, W-w. Lee, Y-S. Jeong, K. Cho, and K. Um. Intuitive NUI for controlling virtual objects based on hand movements. In *Future Information Technology*, volume 309 of *Lecture Notes in Electrical Engineering*, pp. 457–461. Springer, 2014.

P32. G. Drossis, D. Grammenos, C. Birliraki, and C. Stephanidis. Magic: Developing a multimedia gallery supporting mid-air gesture-based interaction and control. In *Proc. of HCI International EA*, pp. 303–307. Springer, 2013.

P33. I. Gatto and F. Pittarello. Prototyping a gestural interface for selecting and buying goods in a public environment. In *Proc. of AVI*, pp. 784–785. ACM, 2012.

P34. E. Hayashi, M. Maas, and Jason Hong. Wave to me: User identification using body lengths and natural gestures. In *Proc. of CHI*, pp. 3453–3462. ACM, 2014.

P35. D. Garzotto and M. Valoriani. Touchless gestural interaction with small displays: A case study. In *Proc. of CHItaly*, pp. 26:1–26:10. ACM, 2013.

P36. S-Y. Lin, C-K. Shie, S-C. Chen, and Y-P. Hung. Airtouch panel: A re-anchorable virtual touch panel. In *Proc. of Multimedia*, pp. 625–628. ACM, 2013.

P37. A. Markussen, M.. Jakobsen, and K. Hornbæk. Selection-based mid-air text entry on large displays. In *Proc. of Interact*, pp. 401–418. Springer, 2013.

P38. F. Alt, S. Schneegass, M. Girgis, and A. Schmidt. Cognitive effects of interactive public display applications. In *Proc. of PerDis*, pp. 13–18. ACM, 2013.

P39. H. Benko and A. Wilson. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *Proc. of ITS*, pp. 19–28. ACM, 2010.

P40 H Benko and A. Wilson. Depthtouch: Using depth-sensing camera to enable freehand interactions on and above the interactive surface. TechReport MSR-TR-2009-23, Microsoft Research, 2009.

P41. M. Karam and M. C. Schraefel. A study on the use of semaphoric gestures to support secondary task interactions. In *Proc. of CHI EA*, pp. 1961–1964. ACM, 2005.

P42. A. Pereira, J. Wachs, K. Park, and D. Rempel. A user-developed 3-D hand gesture set for human-computer interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(4):607–621, 2015.

P43. P. Cremonesi, A. Di Rienzo, and F. Garzotto. Personalized interactive public screens. In *Proc. of Workshop Interacting with Smart Objects at IUI*. ACM, 2015.

P44. E. Freeman, S. Brewster, and V. Lantz. Towards usable and acceptable above-device interactions. In *Proc. of MobileHCI*, pp. 459–464. ACM, 2014.

P45. H. Dong, A. Danesh, N. Figueroa, and A. El Saddik. An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *Access*, 3:543–555, 2015.

P46. J. Shen, Y. Luo, X. Wang, Z. Wu, and M. Zhou. GPU-based realtime hand gesture interaction and rendering for volume datasets using leap motion. In *Proc. of Conference on Cyberworlds (CW)*, pp. 85–92. IEEE, Oct 2014.

P47. S. Stellmach, M. Jüttner, C. Nywelt, J. Schneider, and R. Dachselt. Investigating freehand pan and zoom. In *Mensch & Computer*, pp. 303–312. Oldenbourg Verlag, 2012.

P48. V.K. Adhikarla, P. Wozniak, A. Barsi, D. Singhal, P.T. Kovacs, and T. Balogh. Freehand interaction with large-scale 3D map data. In *Proc. of 3DTV-Conference (3DTV-CON)*, pp. 1–4. IEEE, 2014.

P49. D. Freeman, R. Vennelakanti, and S. Madhvanath. Freehand pose-based gestural interaction: Studies and implications for interface design. In *Proc. of Conference on Intelligent Human Computer Interaction (IHCI)*, pp. 1–6. IEEE, 2012.

P50 S-S. Lee, J. Chae, H. Kim, Y-K Lim, and K-P. Lee. Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In *Proc. of UbiComp*, pp. 617–626. ACM, 2013.

P51. H. Wu and J. Wang. User-defined body gestures for TV-based applications. In *Proc. of International Conference on Digital Home (ICDH)*, pp. 415–420. IEEE, 2012.

P52. B. Vogel, O. Pettersson, A. Kurti, and A. S. Huck. Utilizing Gesture Based Interaction for Supporting Collaborative Explorations of Visualizations in TEL. In *Proc. of Wireless, Mobile and Ubiquitous Technology in Education (WMUTE)*, pp. 177–181. IEEE, 2012.

P53. E. Chan, T. Seyed, W. Stuerzlinger, X-D. Yang, and F. Maurer. User Elicitation on Single-hand Microgestures. In *Proc. of CHI*, pp. 3403–3414. ACM, 2016.

P54. M. Liu, M, Nancel, and D. Vogel. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proc. of UIST*, pp. 63–71. ACM, 2015.

P55. J-W. Lin, C. Wang Y-Y. Huang, K-T Chou, H-Y. Chen, W-L Tseng, and M. Chen. BackHand: Sensing Hand Gestures via Back of the Hand. In *Proc. of UIST*, pp. 557–564. ACM, 2015.

P56. R. Radkowski and C. Stritzke. Interactive Hand Gesture-based Assembly for Augmented Reality Applications. In *Proc. of Advances in Computer-Human Interactions (ACHI)*, pp.303–308. IARIA, 2012

P57. T. Dias, M. Variz, P. Jorge, and R. Jesus. Gesture Interaction System for Social Web Applications on Smart TVs. In *Proc. of Open Research Areas in Information Retrieval (OAIR)*, pp. 225–226. Le Centre de Hautes Etudes Internationales D'Informatique Documentarie, 2013.

P58. R-D. Vatavu and I-A. Zaiti. Leap Gestures for TV: Insights from an Elicitation Study. In *Proc. of Interactive Experiences for TV and Online Video (TVX)*, pp. 131–138. ACM, 2014.

P59. J. Yim, K. Lee, and H. Kim. Implementation of a Preliminary Natural User Interface for Video on Demand Systems. In *Future Information Technology: FutureTech*, pp. 475–480. Springer, 2014.

P60. R. Salvador, T. Romão, and P. Centieiro. A Gesture Interface Game for Energy Consumption Awareness. In *Proc. of Advances in Computer Entertainment (ACE)*, pp. 352–367. Springer, 2012.

P61. A. Löcken, T. Hesselmann, M. Pielot, N. Henze, and S. Boll. User-centered approach for the definition of free-hand gestures applied to controlling music playback. *Multimedia Syst.* 18(1): pp. 15–31. Springer, 2012.

P62. I-A. Zaiţi, Ş-G. Pentiuc, and R-D. Vatavu. On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. *Personal and Ubiquitous Computing*, 19(5): pp. 821–838, August, 2015.

P63. T. Sousa, I. Cardoso, J. Parracho, P. Dias, and B. S. Santos. DETI-Interact: Interaction with Large Displays in Public Spaces Using Kinect. In *Proc. of Distributed, Ambient, and Pervasive Interactions (DAPI)*, pp. 196–206. Springer, 2014.

P64. S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin. Instructing people for training gestural interactive systems. In *Proc. of CHI*, pp. 1737–1746. ACM, 2012.

P65. X. Yi, C. Yu, M. Zhang, S. Gao, K. Sun, and Y. Shi. ATK: Enabling Ten-Finger Freehand Typing in Air Based on 3D Hand Tracking Data. In *Proc. of UIST*, pp. 539–548. ACM, 2015.

## REFERENCES

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

Biolchini, J., P. Mian, A. Natali, T. Conte, and G. Travassos (2007, April). Scientific research ontology to support systematic review in software engineering. *Adv. Eng. Inform. 21*(2), 133–151.

Bowman, D., E. Kruijff, J. La Viola, and I. Poupyrev (2004). *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing.

Chan, E., T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer (2016). User elicitation on single-hand microgestures. In *Proc. of CHI*, pp. 3403–3414. ACM.

Erazo, O. and J. Pino (2015). Predicting task execution time on natural user interfaces based on touchless hand gestures. In *Proc. of IUI*, pp. 97–109. ACM.

Giustini, D. and M. Boulos (2013). Google scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics 5*(2).

Gough, D., S. Oliver, and J. Thomas (2013). *An Introduction to Systematic Reviews*. Sage Publications.

Karam, M. and M. C. Schraefel (2005). A taxonomy of gestures in human computer interactions. Technical Report 261149, University of Southhampton.

Kelley, J. F. (1984, January). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst. 2*(1), 26–41.

Kitchenham, B., o. P. Breretona, D. Budgen, M. Turner, J. Bailey, and S. Linkman (2009, January). Systematic literature reviews in software engineering - a systematic literature review. *Journal of Information and Software Technology 51*(1), 7–15.

Kitchenham, B. and S. Charters (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Keele University.

Liu, J. and M. Kavakli (2010, July). A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games. In *Proc. of Conference on Multimedia and Expo (ICME)*, pp. 1564–1569. IEEE.

Piumsomboon, T., A. Clark, M. Billinghurst, and A. Cockburn (2013). User-defined gestures for augmented reality. In *Proc. of Interact*, pp. 282–299. Springer.

Ruiz, J., Y. Li, and E. Lank (2011). User-defined motion gestures for mobile interaction. In *Proc. of CHI*, pp. 197–206. ACM.

Seyed, T., C. Burns, M. Costa Sousa, F. Maurer, and A. Tang (2012). Eliciting usable gestures for multi-display environments. In *Proc. of ITS*, pp. 41–50. ACM.

Shakeri, Z., C. Anslow, and F. Maurer (2014). Multi surface interactions with geospatial data: A systematic review. In *Proc. of ITS*, pp. 69–78. ACM.

Sowa, T. (2008). *Modeling Communication with Robots and Virtual Humans*, Chapter The Recognition and Comprehension of Hand Gestures - A Review and Research Agenda, pp. 38–56. Springer.

Strauss, A. and J. M. Corbin (1997). *Grounded theory in practice.* Sage.

Vatavu, R.-D. (2012). User-defined gestures for free-hand tv control. In *Proc. of European Conference on Interactive Tv and Video (EuroiTV)*, pp. 45–48. ACM.

Wobbrock, J., M. Morris, and A. Wilson (2009). User-defined gestures for surface computing. In *Proc. of CHI*, pp. 1083–1092. ACM.

Wolfswinkel, J., E. Furtmueller, and C. Wilderom (2013). Using grounded theory as a method for rigorously reviewing literature. *Eur J Inf Syst 22*(1), 45–55.

Wu, Y. and T. S. Huang (1999). Vision-based gesture recognition: A review. In *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, pp. 103–115. Springer-Verlag.