



Systematic literature review of hand gestures used in human computer interaction interfaces



Tijana Vuletic*, Alex Duffy, Laura Hay, Chris McTeague, Gerard Campbell, Madeleine Grealy

Design Manufacturing Engineering Management, University of Strathclyde, 75 Montrose Street, DMEM, 7th Floor, Glasgow G1 1XJ, United Kingdom

ABSTRACT

Gestures, widely accepted as a humans' natural mode of interaction with their surroundings, have been considered for use in human-computer based interfaces since the early 1980s. They have been explored and implemented, with a range of success and maturity levels, in a variety of fields, facilitated by a multitude of technologies. Underpinning gesture theory however focuses on gestures performed simultaneously with speech, and majority of gesture based interfaces are supported by other modes of interaction. This article reports the results of a systematic review undertaken to identify characteristics of touchless/in-air hand gestures used in interaction interfaces. 148 articles were reviewed reporting on gesture-based interaction interfaces, identified through searching engineering and science databases (Engineering Village, Pro Quest, Science Direct, Scopus and Web of Science). The goal of the review was to map the field of gesture-based interfaces, investigate the patterns in gesture use, and identify common combinations of gestures for different combinations of applications and technologies. From the review, the community seems disparate with little evidence of building upon prior work and a fundamental framework of gesture-based interaction is not evident. However, the findings can help inform future developments and provide valuable information about the benefits and drawbacks of different approaches. It was further found that the nature and appropriateness of gestures used was not a primary factor in gesture elicitation when designing gesture based systems, and that ease of technology implementation often took precedence.

1. Introduction

Gestures are a natural form of human expression, and hands a natural mode of interaction with the physical world and objects in it (Zimmerman et al., 1987; Buchmann et al., 2004). Gestures are used for communication and accompany speech in many different forms. They range from gestures that do not convey a specific meaning and simply follow the rhythm of the speech, to those enriching its meaning and symbolizing specific concepts (McNeill, 1992; Quek, 2004).

Hand gestures in particular, including use of fingers and arms, are widely explored as a natural and intuitive interaction modality for a variety of applications. They are used as a sole, or one of the modes for interaction interfaces. It is believed that gesture based interfaces can reduce the complexity of interaction between humans and computers (New et al., 2003). Motivations behind the decision to use gestures in an interface can be varied. Gesture based interfaces used for computer applications can be more intuitive than established WIMP (Windows Icon Mouse Pointer) based interfaces, and allow inexperienced users to interact with computer applications, without undertaking extensive training (Buchmann et al., 2004; Kim et al., 2005; Beyer and Meier, 2011). In medical applications or industrial environments, they enable touchless operation guaranteeing sterility or safer interaction. Gesture interfaces for Virtual Reality (VR) or Augmented Reality (AR)

environments provide better immersion and do not require conscious attention dedicated to the specific gestures being performed (Deller et al., 2006). Spatial concepts can be expressed using gestures, and they are used in design and engineering, when externalising ideas (Vinayak et al., 2013). Interaction with comfort functions in a car can be achieved without taking the eyes off the road (Riener et al., 2013). Gestures can be used to help older population achieve easier interaction with electronic devices (Bhuiyan and Picking, 2011). These are just some of the examples, and new applications are constantly being developed. Use of gestures for these applications is supported by a variety of technologies. Development of Kinect (Kinect, 2018) and LEAP sensors (LEAP MOTION INC., 2018), which are portable and supported by Software Development Kits (SDKs) enabling simpler implementation, seems to have contributed significantly to the expansion of the field on gesture based interfaces since 2013.

While the gesture-based interfaces are being developed for various applications clear standards which could guide their further development are not apparent. For example, while interfaces supporting three-dimensional (3D) object manipulation exploring use of intuitive, affordable and non-intrusive interfaces are ubiquitous, none of the approaches used have been established as the baseline for future development (Vinayak et al., 2013). Investigation of patterns of gesture use, identifying commonalities and differences between different fields

* Corresponding author

E-mail address: tijana.vuletic@strath.ac.uk (T. Vuletic).

<https://doi.org/10.1016/j.ijhcs.2019.03.011>

Received 25 September 2018; Accepted 29 March 2019

Available online 30 March 2019

1071-5819/ © 2019 Elsevier Ltd. All rights reserved.

would be an initial step towards development of a standard framework for gesture elicitation for interaction interface development.

Review by [Rautaray and Agrawal \(2015\)](#), provides a survey of the gesture based research published up to and including 2012, and the content covered is largely that published prior to the uptake of Kinect and LEAP in the gesture research community. Reviews by [Hasan and Mishra \(2012\)](#), [Suarez and Murphy \(2012\)](#), and [Pisharady and Saerbeck \(2015\)](#) focus on recognition approaches. Three systematic reviews have been identified: one that focuses on usability guidelines for “health serious games” ([Milani et al., 2017](#)), one that focuses on data exchange formats ([Santos¹ et al., 2015](#)), and one that focuses on vision based gesture systems and algorithms for gesture recognition ([Al-Shamayleh et al., 2018](#)). The first one is in Portuguese and reports on only 16 studies. Reviews identified in the literature either cover recognition based research questions, rather than gestures themselves, or are not systematic, and information on patterns of gesture use cannot be extracted from them.

Further to this, the underpinning gesture theory found in the literature is heavily based on gestures observed as speech aid, gestures used in parallel with verbal communication. Classifications and definitions present in it may not be capable of fully describing the free-form in-air gesture interaction often used for interaction with 3D/VR/AR environments for example, or be aligned with its goals. This will be explored further in [Section 2.2](#).

This paper reports findings from a systematic review aiming to answer the following research question: What are the patterns of touchless hand gesture use during gesture based interface design for interfaces which have reached the prototype stage? To answer this research question, the field of touchless gesture based interfaces is first mapped, and then patterns and commonalities between different approaches are explored. The nature of gestures used in different interaction interfaces is explored, along with the reasoning, if present, behind the gesture elicitation choices made in them. The nature of gestures is characterised by the motions used and the role they serve in the interface. The impact of the application field, technology, and recognition techniques on gesture use is also observed. The aim was to identify and classify types of gestures used in interaction interfaces, and explore similarities and differences encountered in their elicitation and implementation. While one of the inclusion criteria is the existence of a prototype, the interest was in the nature and type of gestures used in these prototypes, rather than implementation details. They are analysed in terms of the role the gestures used play, and classified based on established gesture theory. Admittedly, the goals of work focusing on application of gestures in an interface and theoretical classification of gestures based on their role in a communicative process are fundamentally different. However, articles reporting on prototypes of gesture-based interfaces often classify gestures using gesture theory, and exploring how well they can describe the gestures used was deemed worthwhile.

[Section 2](#) provides the context and history of gesture research, and establishes the terminology and the classification approach that will be followed in this paper. [Section 3](#) provides a description of the methodology followed in the article. [Section 4](#) classifies the types of gestures identified in the articles reviewed. [Section 5](#) summarizes application types encountered, and technologies and recognition approaches that facilitate them. Analysis of the role and nature of gestures used in different applications is reviewed in [Section 6](#), and patterns of gesture use are covered in [Section 7](#). Discussion in [Section 8](#) establishes links between the gestures and applications, technologies and recognition approaches, and discusses the potential paths for future research.

2. Gestures and computer interaction – definitions

In this article hand gestures are defined as gestures performed using one or both hands, including finger gestures when they were performed along with a number of other varied gestures e.g. pointing gesture is

used for selection of an object and then pinching gesture is used to deform that object, or to move it to a different location. If only fingers were used to trigger an action then gestures were classified as finger gestures e.g. touching different combinations of fingers together triggers an action, where that is the only type of gesture accepted by a system. In some articles, only one hand was used to perform the gestures. This was particularly prominent in the older articles, and it was usually caused by technical inability to track both hands due to lower bandwidth or processing power of devices used. Since two handed gesture recognition no longer poses a technical issue, due to technology developments, distinction between one-handed and two-handed gestures was not made during the analysis, unless it was significant for some other, non-technical reason. Handedness was also not observed as it does not fundamentally change the type of gestures observed i.e. the role a gesture plays in the communicative process with the computer. For example, [Kang et al. \(2013\)](#) developed an interface where the posture of the left hand defined the operation that the designer wanted to perform, and right hand performed a dynamic action specifying parameters of it. On the other hand, [Bourdout et al. \(2010\)](#) explored manipulation of the 3D object with one hand, while slave hand could be used to control the 3D scene. This corresponds to the way handedness is observed in everyday life, to an extent. Some gestures are unimanual (e.g. brushing teeth or throwing darts) some bimanual symmetric (lifting weights or rope skipping), but vast majority are bimanual asymmetric (dealing cards or playing violin) ([Guiard, 1987](#)). Usually two hands cooperate and form a series of actions building on each previous one ([Guiard, 1987](#)). Within bimanual asymmetric gestures, temporal and spatial scales influence cooperation between two hands. For example while writing with one hand and holding and occasionally adjusting the paper with the other, hand which is holding the paper has a low temporal frequency (periods between changes are longer), and a low a spatial frequency (movement itself is larger in comparison to the writing hand) ([Guiard, 1987](#)).

Gestures have been classified based on their temporal characteristics, contextual characteristics and the levels of instruction present.

2.1. Temporal classification

Based on their temporal characteristics, gestures can be classified as static or dynamic. Static gestures are those where only static position of a hand (“a snapshot”) is observed. These types of gestures are also called postures ([Vinayak et al., 2013](#)). Dynamic gestures are those in which a hand moves between a number of positions to form a full gesture. A dynamic gesture is considered to have three to five phases, depending on the specific gesture performed: preparation, prestrike hold (if specific gesture contains it), stroke, post stroke hold (if specific gesture contains it), and retraction ([McNeill, 1992](#)).

2.2. Contextual classification

Based on what they are used for, and the context they are used in, gestures can be classified as communicative or manipulative ([Quek, 1995](#)). It can be argued that even manipulative gestures communicate spatial position of objects, or the ways they are manipulated, but for clarity communicative gestures referred to in this paper will encompass gestures that have a link to speech. [Quek \(2004\)](#) states that “gesture and speech are complementary, and have developed as equal communication features”. Different authors have different approaches to classification of communicative gestures into different gesture types. Summary of these classifications is shown in [Table 1](#). In it gesture types are further classified concerning the similarity their structure bears to language (based on work by [Kendon \(1988\)](#) as cited in [Billinghurst \(2018\)](#)), if they can be used independently from speech, what they communicate (semiotic – communicate meaningful information, ergotic – manipulate the physical world to create artifacts, or epistemic – learn from environment through tactile or haptic exploration; based on work

Table 1
Communicative gesture classification.

Do gestures have meaning independent of speech	Link to language	Gesture type	What does a gesture communicate (function)	What does a gesture refer to	Conscious or unconscious
Independent	Language (of its own)	Symbolic (Rimé and Schiaratura, 1991; Quek, 1995)/ Emblematic (Wagner et al., 2014) Semaphoric (Quek, 2004; Santos et al., 2016) Pantomimic (Rimé and Schiaratura, 1991; Quek, 1995; Boulabiar et al., 2011) Iconic (McNeill, 1985; McNeill, 1987) Metaphoric (McNeill, 1985) (also Ideographs (Rimé and Schiaratura, 1991)) Modalizing symbolic (Quek, 1995)/Speech linked (McNeill, 2006)/Speech marking (Rimé and Schiaratura, 1991)/Beat (Wagner et al., 2014)	Semiotic (Communicate meaningful information)	Object (evocative)	Conscious
Speech related	Pantomimes (medium link to language) Language like (some link to language) Gesticulation (low link to language)			Object of speech (depictive) Ideation process (depictive) Ideation process (nondepictive)	Unconscious
	Emblems (High link to language)	Cohesive (Rautaray and Agrawal, 2015)/Catchments (Quek, 2004) Adaptors (Rautaray and Agrawal, 2015)/Butterworths (McNeill, 2006) Deictic (McNeill, 1987; Rimé and Schiaratura, 1991; Quek, 2004)		Object (evocative)	Conscious (when interacting with an object) / Unconscious (when part of a sentence)

by Cadoz (1994)), what do they refer to (evocative object, object of speech or an ideation process; based on work by Rimé and Schiaratura (1991)) and if they are conscious or not (based on work by Cassell (1998)). In order to form a basis for pattern discovery during the literature review the interest is in exploration of how well the established gesture theory classified in Table 1 describes the practical applications of gestures in interfaces. This, particularly gesture types and their link to language, is explored in Section 6.

The first group of gestures, **independent communicative gestures**, are those that whilst still having a communicative purpose and most frequently are used concurrently with speech, can be independent from speech, and do not require it to convey a meaning. **Symbolic gestures** represent a symbolic object or concept (Quek, 1995), devoid of any morphological relation with what is being referred to (Rimé and Schiaratura, 1991). They have a direct translation into words, are used deliberately to send a particular message (Rimé and Schiaratura, 1991), and have a widely accepted meaning (Wagner et al., 2014), albeit one that may be specific to a group, class or culture. Some examples are “thumbs up” to indicate approval (Wagner et al., 2014), hand waving as a greeting (Rimé and Schiaratura, 1991) or “rubbing index finger and the thumb to refer to money” (Quek, 1995). (Wagner et al., 2014) classify them as emblematic. **Semaphoric** gestures are used to trigger a predefined action, defined in a formalised dictionary and therefore “require prior knowledge and learning” (Santos et al., 2016). They are “static poses or predefined stylized movements communicating an intended symbol to a machine” (Quek, 2004). They are similar to symbolic gestures, in terms that the meaning is predefined and that both refer to an object (are evocative), but the semaphoric gesture lack the potential common background, and are instead developed for a specific purpose. This may also mean that semaphoric gestures may potentially be used to communicate a wider range of information, not just the verbal information the symbolic gestures focus on. **Pantomimic gestures** represent familiar concepts, but they are imitations of what is being implied e.g., “motioning ‘lighting up’ of a cigarette to ask for a lighter” (Quek, 1995). Pantomime is also defined as “a sequence of gestures conveying a narrative, produced without speech” (Boulabiar et al., 2011). Pantomime “plays the role of the referent”, for example if a speaker talks about a person grasping a box, they would shape an imaginary box with their hands at the same time (Rimé and Schiaratura, 1991).

The second group of gestures are **speech related** and complement what is being communicated verbally. **Iconic gestures** “represent meaning closely related to the semantic content of the speech” (Holler and Beattie, 2003; McNeill, 1985) and illustrate what is being said. They “present some figural representation of the object evoked in speech” (Rimé and Schiaratura, 1991). For example, a person discussing an object rolling down a hill would perform a rolling motion using their hands. There are three sub-types of iconic gestures: those that describe a shape (Pictographs), those that represent a spatial relation (Spatiographic), and those that describe action of an object (Kinematographs) (Rimé and Schiaratura, 1991). **Metaphoric gestures** “are iconic gestures which represent abstract content” (Wagner et al., 2014; McNeill, 1992), e.g. a cutting gesture to indicate a decision has been made (Casasanto and Lozano, 2007). They “sketch in space the logical track followed by the speaker's thinking” (Rimé and Schiaratura, 1991). **Modalizing symbolic gestures** primarily complement speech, but can also complement other means of communication. For example, a person asking “Have you seen her husband?” while holding their hands apart would indicate he is overweight” (Quek, 1995). Other authors call these speech-linked gestures or speech framed gestures (McNeill, 2006), as they most frequently complete spoken statements (Boulabiar et al., 2011). Rimé and Schiaratura (1991) call them speech marking gestures. These gestures stress some elements of speech to clarify them, usually coincide with the introduction of some new element on the discourse, and chunk the sentence to highlight its underlying reasoning. Wagner et al. (2014) call them beat gestures, and

define them as “simple and fast movements of hand that synchronize with prosodic events, variations in pitch, loudness, tempo, and rhythm, of speech”. **Cohesive gestures** are “those that are thematically related, but temporally separated”, where a continuation of a specific theme after the speaker was interrupted is characterised by the recurrence of a gesture (Rautaray and Agrawal, 2015). These recurring gestures are perceived to be related to the recurring themes contained in the speech, and are in other sources referred to as catchments (Quek, 2004; Yoshioka, 2005). ‘**Butterworth’s**’ were thought to be gestures that “arise as failures of speech e.g. hand grasping while a speaker is trying to recall a word” (McNeill, 1992). However, researchers have been unable to replicate the claimed frequency of these gestures in later studies, and Butterworth’s gestures are now thought to be misrepresented (McNeill, 2006). **Adaptors** “are gestures like headshaking or quickly moving one’s leg that are unconscious and used to release body tension” (Rautaray and Agrawal, 2015).

Deictic gestures are pointing gestures, used to indicate the direction of intended movement, or a direction of manipulation. Depending on the context and the direction of pointing, they can also have assigned meaning. They are similar to manipulative gestures, or integrated with the manipulative gestures, but they are more limited (Quek, 2004). They can point towards some visually or symbolically present object, which may be a place of an event (Rimé and Schiaratura, 1991). Some authors argue they might represent an abstract form of iconic gestures (McNeill, 1987), and be one of the first gestures children are capable of performing.

Manipulative gesture classification is far less extensive. Gestures are generally considered manipulative if they interact with and modify a spatial component of an object in an interface. Quek (2004) states that in manipulative gestures “hand motion indicates a path or a placement”. They are classified in Table 2, following the same convention as the one followed in Table 1.

Deictic gestures could be classified as a mix of manipulative and communicative gestures, due to their directional nature.

Gesticulation is a term used to designate gestures performed simultaneously with speech; meaning that the stroke component of the gesture is performed at the same time as related verbal content is spoken, and it has been found that they accompany speech in that manner 90% of the time (McNeill, 2006). Gestures help get the meaning across by supplementing speech, substituting speech, completing a sentence or clarifying ambiguous words (Kendon, 1985, pp. 215–234). McNeill (1987) believes thinking is image based and speaking is syntactic, and suggests that concurrent use of speech and gestures means that a person is engaging simultaneously in both types of thinking. Church and Goldin-Meadow (1986) find that gestures attempt to identify underlying reasoning processes that the speaker did not or could not articulate. For example, when mental image, for which the meaning is independent of language formats (McNeill, 1987), cannot be easily translated into words, the speaker would use gestures in addition to speech (Freedman, 1972). Kendon, on the other hand, focuses on anthropology and centres on origins of language. He believes what is spoken is structured according to the rules of the language system, and only indirectly related to any aspect of the structure that is being referred to, which means that gestures, which do not have rules to follow have more degrees of freedom of expression (Kendon, 1986). This may indicate that further research into ergotic and epistemic gestures, focusing on mental images independent from language formats is needed.

This will be further discussed in Section 8.3.

Discussion on typification and classification of gestures, while extensive and present in the field since the 1940s, focuses on gestures that are in relation to and performed concurrently with speech. This systematic review focuses on the gestures used in the applications which reached the implementation stage, and in them gestures are not always used along with speech and they do not always convey activity that is easily put into words. Section 6 explores how gestures implemented in the applications map onto the gesture focusing on verbal communication. Gestures found in the literature are classified as communicative (specific types of communicative gestures from Table 1) or manipulative in Section 6, and then their link to speech, if it exists, and the form it is in is explored.

2.3. Instruction based classification

Based on levels of instruction given to guide the gesture performance, gestures can be classified as prescribed or free-form gestures. Prescribed gestures are those in which a dictionary of gestures is defined prior to being used, the users of an application have to learn these gestures, and the performance of a predefined gesture triggers a predefined action. Prescribed gestures may increase the cognitive load (Poupyrev et al., 2002; Shapir et al., 2007), their learning rates depend on the users’ cognitive skills (Wachs et al., 2011), and their use forces the users to learn and use gestures they perhaps would not choose themselves. Free-form gestures are unrestricted and typically do not trigger specific uniform predefined actions. In the context of interactive interfaces, they are normally copied into the system the interfaces are used for, and are often used to form splines or surfaces, or move objects in a virtual space. This means they do not communicate symbolic or metaphoric meanings which prescribed gestures can convey. Free-form gestures therefore, contrary to the implication of being non-restrictive that their name infers, have a limited application breadth in their current form.

3. Methodology

Prior to conducting a systematic review, a background review focusing on gestures and their definitions was performed. This review provided the conceptual framework and guided the interpretation of the systematic review findings. The background review findings were reported in Section 2.

The nature of the topic explored is qualitative, and studies included span a large time frame and include diverse methodologies. Due to this, a systematic review following the qualitative synthesis approach was performed. The approach was informed by the PRISMA guidelines (Moher et al., 2009) applicable to qualitative synthesis.

3.1. Article selection process

Engineering and science databases searched were: Engineering Village, Pro Quest, Science Direct, Scopus and Web of Science.

First search was completed on Feb 22nd 2016 and contained following words:

("gesture interaction" OR "gesture interactions" OR "gestures interaction" OR "gestures interactions")

Table 2
Manipulative gesture classification.

Do gestures have meaning independent of speech	Link to language	Gesture type	What does a gesture communicate (function)	What does a gesture refer to	Conscious or unconscious
Independent	Low link to language	Manipulative	Ergotic gestures – manipulate the physical world to create artifacts	Object of speech (depictive)/ Ideation process (depictive)	Conscious

AND interface AND control AND human AND computer.

As this search focused on gestures and interaction only, it was evident first through abstract and then following the full text review, that some of the papers authors were aware of using gestures for CAD application were missing. Therefore, a second search was performed focusing on 3D modelling and CAD applications that resulted in a prototype or an implementation, on October 6th 2016 using following keywords:

gesture AND "cad" AND engineering AND interface AND control AND human AND computer AND design AND 3D AND hand AND (prototype OR simulation OR application).

The two searches combined resulted in 545 papers. (262 in search one, and 283 in search two). After the removal of duplicates 488 papers remained. Of these, nine papers were not accessible (six in search one and three in search two). Two papers found in the first search were in Chinese, although the search was limited to papers in English. After the abstract screening 183 remained (141 in search one, and 42 in search two). After the full paper review, 92 papers were deemed applicable to the research questions (74 in search one, and 18 in search two).

Reference search of the included articles was performed. As the titles generally did not contain all the keywords, the criteria adopted was that a paper was considered if the title indicated gestures were used and there was a practical application i.e. a prototype or an experiment. Following the abstract and then full paper review of the 70 papers selected based on the title, 33 articles were included in the review.

Another search was performed on August 29th 2018 to account for any papers published since, using both sets of keywords specified earlier, and out of 61 identified papers, 23 were added from this search.

In total 148 papers were included that tested a specific prototype of a gesture interface.

Fig. 1 illustrates the stages of the systematic literature review, suggested by the (Moher et al., 2009).

Articles were included if they primarily explored touchless hand gestures and reported on implementation of a gesture interface, through experimental testing or prototype evaluation. If interaction was aided with hand held devices such as wands or remotes, they were required to participate in gesture formation i.e. users did not simply point the devices and press buttons, or write using a wand. No full body gestures were included, but if arm or upper body were used along the hand gestures, they were included. Touchless gestures were the focus of this article, hence tablet, touch and pen applications were not included.

3.2. Sample characteristics

More than a third of the articles (53 out of 148) do not report on the number of participants included in the study. The most common number of participants was 10–15 participants, and 36 articles fall into that range. 26 articles included less than 10 participants, 27 articles included 15–30 participants, one article had 32 participants, and two articles had 37 participants each. Articles that included significantly larger numbers of participants each had 40, 67, 70, and 100 participants. Carreira et al. (2017) evaluated in-air hand gesture interaction for older people, using prescribed swipe, grab and drag, point and hold, and point and push gestures performed by 40 users in two different countries. He et al. (2008) tested 17 predefined gestures performed while holding a mobile phone in hand, using 67 participants. Bhuiyan and Picking (2011) used 70 participants to test a gesture based system enabling elderly users to control a TV, however specific gestures used were not reported. The study which included 100 participants asked them to hold a phone and perform a single free-form gesture to unlock it (Guerra-casanova et al., 2012). These articles did not classify or compare the gestures created by different participants, and while the larger number of participants was significant for the purposes of machine learning and gesture recognition mechanisms being developed, it

was not influential in terms of observing the patterns of gesture elicitation. Only one article reports on a study in which participants were asked to use any gestures they wished for predefined tasks, that were then taught to the system (Kim et al., 2017). Some articles report on 61 video sequences analysed or 120 samples, which does not necessarily mean participants, and number of participants is not stated. Some articles test slightly different versions of their systems more than once, with different number of participants, and they have been counted as different applications.

82 out of 148 reviewed articles tested six or less gestures, and while this is not necessarily a drawback, it might indicate that gestures were used for limited and specific forms of interaction. The exception are free-form gestures, which due to their nature had included an unlimited number of gestures, but only six papers tested free-form gestures.

While it is difficult to provide general guidance for indicative number of participants to ensure meaningful outcomes, one suggestion would be that the number of participants must allow for statistical analysis that results in significant findings. This tends to be more likely with participant numbers higher than 20, but preferably even higher than that. While whether there is correlation between number of gestures used and number of participants testing them was not explored in this review, number of gestures would not inherently influence the statistical analysis. However, where gestures are explored in the context of an application aiming to gain widespread adoption, the number of participants testing and evaluating the implementation outcomes would become more important.

Sample size in terms of participant numbers and number of gestures tested are shown in Fig. 2.

A typical article reporting on a gesture interface implementation first reviews the literature on similar applications, reports on technology, gestures and conditions used, discusses recognition methods and then reports on the findings. Ten papers did not provide information on the gestures used, or specific activities which were performed, and do not provide enough information for a study to be considered fully defined and repeatable. They were included in the review, as their findings were still valuable inclusion during the mapping of the field. However, these were considered to be a lower quality papers, and are noted as such in the table listing the articles, given in the appendix. For example, Dave et al. (2013) claim to use a software which allows 3D painting and low level gestures to manipulate the models, however not enough detail is given in the paper to confirm this.

3.3. Qualitative synthesis

For the included articles, the following information was mapped:

- What gestures were used?
- What type of gestures were they (based on temporal characteristics, context they are used in and rules guiding them e.g. static/dynamic, symbolic/deictic/iconic/etc., prescribed/free-form)?
- What were they used for, how were they recorded and tracked?

Then the gestures were classified based on definitions present in the field of gesture theory research (in this article referred to as contextual classification). However, when gestures reach the application stage this means that two gestures, while technically performing the same motion, can have completely different meaning and role in the context they are used in, and be classified differently. For example, open hand moving upwards used to pan a 3D model vertically upwards in a virtual space would be classified as a manipulative gesture, and the same gesture used to turn a television screen on as a part of a home appliance control interface would be classified as a semaphoric gesture. Rules for coding the gestures identified in the reviewed articles, depending on their role in the application, were given in Table 3.

Some studies used only one gesture type, and some used multiple types within the same solution. After gestures were classified based on

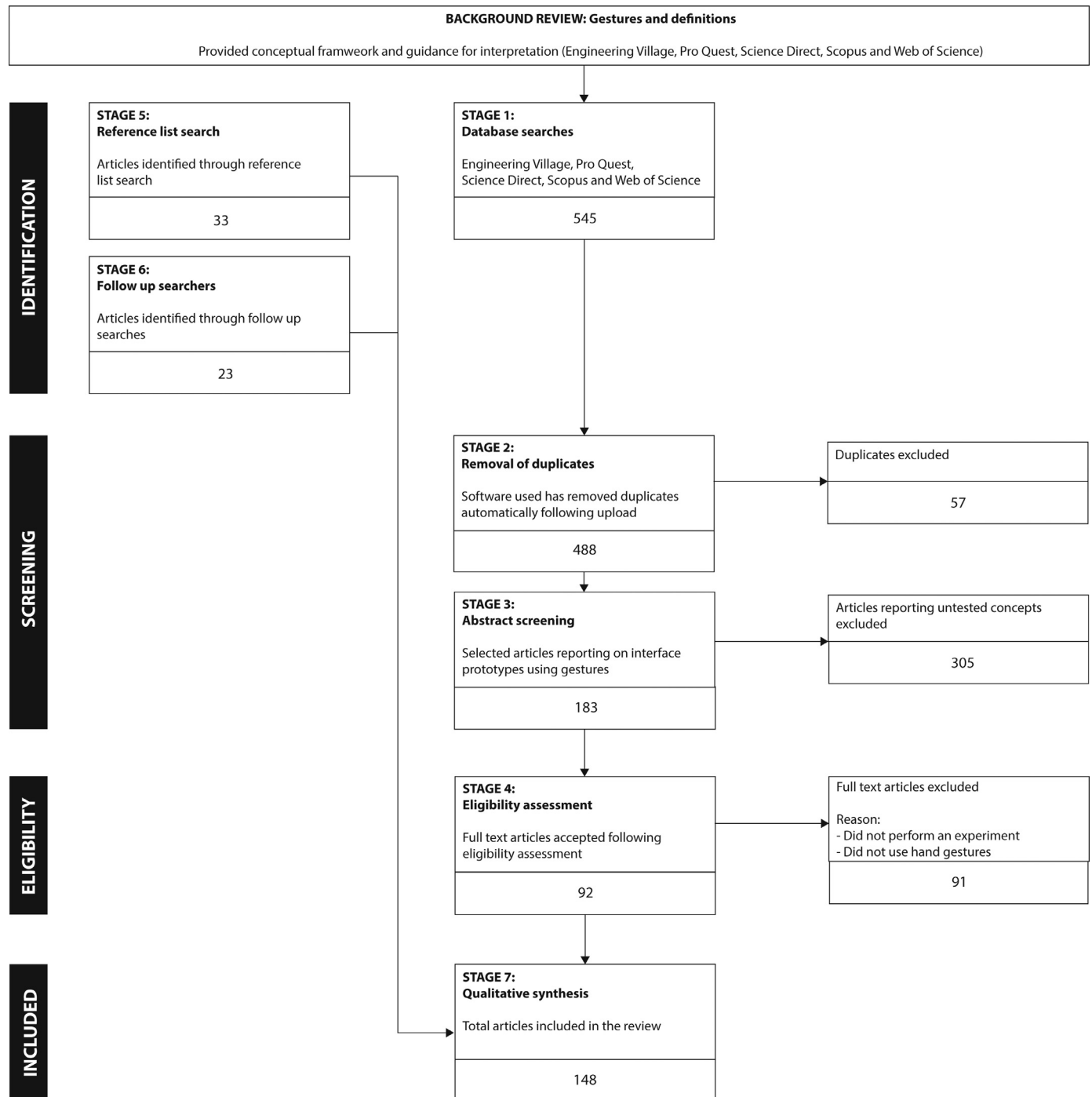


Fig. 1. Systematic review process.

their role, and the context they were used in, a higher level view was taken, observing the contextual nature of the gesture, technology used to implement it and the application field, and patterns were observed. This was achieved by graphically representing and grouping the articles using a Venn diagram. In total, eleven distinct combinations of contextual gesture types were identified, used across applications and supported by different technologies, and three high level combinations based on the technology supporting the gestures were identified. Both are detailed in Section 6.

4. Types of gestures

In this section gestures observed in the reviewed articles were

classified based on the terms and definitions given in Section 2. First, they were classified based on how well they fit the definition of a hand gesture. Then the classification was performed by reviewing the examples of specific gestures given in the articles, and assigning them a class based on temporality, context and levels of instruction guiding them.

Classification of articles based on adherence to the given definition of a hand gesture is shown in Fig. 3. The order of gestures was based on size of gestures, from smallest to largest. In some articles authors discussed hand gestures, although they really focus on finger or arm gestures only, and this has been accounted for in Fig. 3. While these gestures did not fully fit the definition stated in this article, they were still included in the review as the way they were used was very similar to

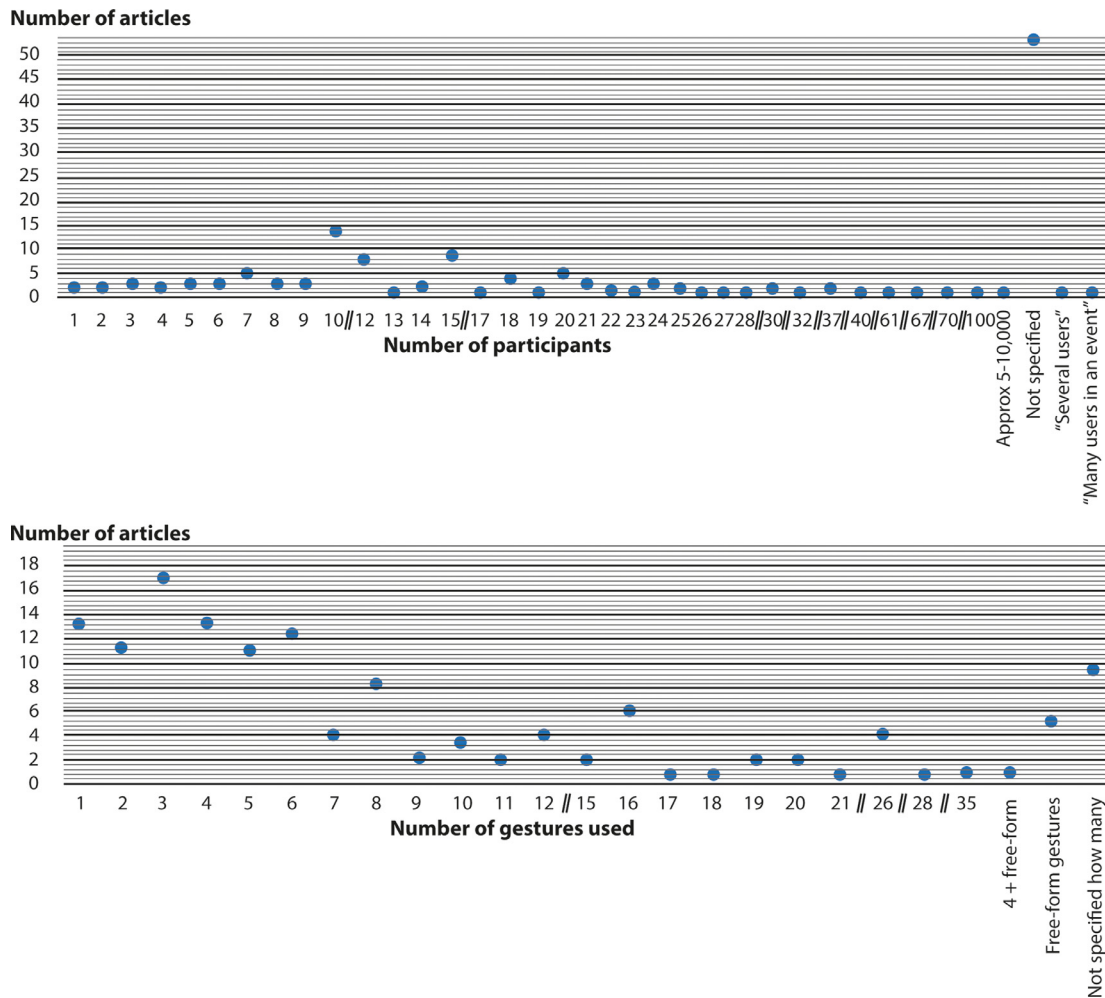


Fig. 2. Number of participants and gestures used identified in the sample.

the way hand gestures encountered were used. In some articles the examples of gestures were not given and gestures were not described. In those cases, it was impossible to confirm if they were actually hand gestures, and they were classified as “Not specified”.

Hand gestures were observed in 136 articles (92%) included in this review. In Fig. 3, one box corresponds to one article. Nine articles contain only gestures which would technically be classified as finger, arm, or upper body gestures. Three articles do not provide enough information to be classified.

4.1. Static and dynamic gestures (temporal view)

In the majority of the articles (139 articles) gestures were dynamic. In two articles majority of gestures were dynamic, with added static gestures used for selection. In these two cases gestures used in an article were classified as dynamic, since the interface used would require the capability to recognize both types of gestures. Four interfaces employed only static gestures, and three did not define the gestures used as static or dynamic, or provide sufficiently clear description of gestures that would have enabled their categorisation. Static and dynamic gestures are also represented in Fig. 3, dynamic gestures are the dark boxes, static gestures are medium shaded boxes, and where light boxes

Table 3
Coding for contextual gesture types.

Activity performed with a hand	Aim/role	Classification
Rhythmic gesture	Gesture where shape is not traced, but timing	Modalizing symbolic gestures/ Beat gestures
Pointing with one, two fingers or a fist	To indicate direction, selection or location of a single point	Deictic (Point to select)
Gestures emulating interaction with a physical object, e.g. pressing a button	To continually move an object or indicate a path	Deictic (Point to move)
Gestures vary, but are often one finger, two fingers of a full hand moving in a certain direction or tracing a circle	Same goal and role as with the emulated physical object, e.g. button is pressed	Pantomimic
Gesture is predefined, needs to be performed accurately and does not have to have any logical tie to the event it triggers	To physically manipulate something in an interface, e.g. translate it, rotate it, scale it, or trace an object shape	Manipulative
No limitations, hand moves freely	To trigger a predefined event	Semaphoric
	Reproduction of the motion in the system	Free-form (Reproduction)
	Virtual hand emulates movement	Free-form (Virtual hand)

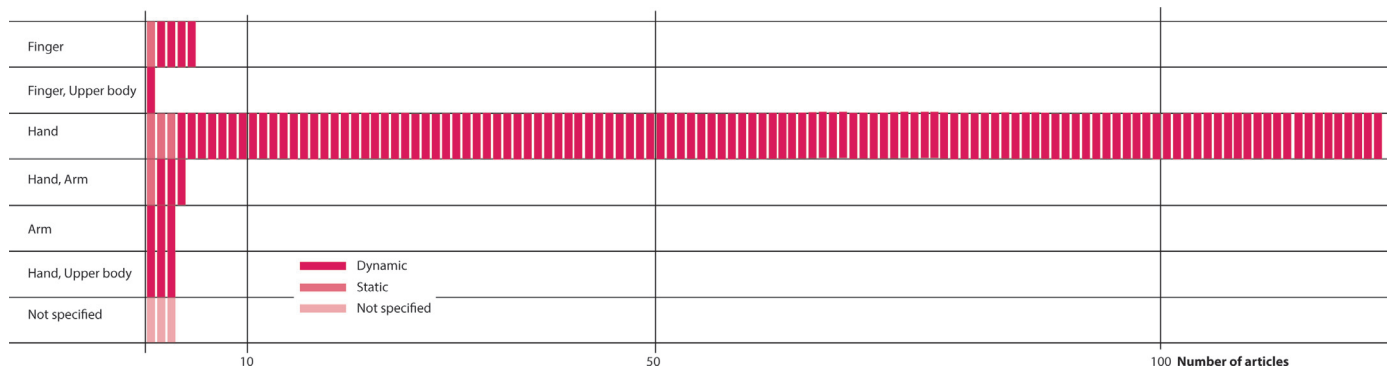


Fig. 3. Finger, hand, arm and upper body gestures.

represent the articles with insufficient information.

4.2. Communicative or manipulative gestures (contextual view)

In multimodal applications, speech and gestures were often combined (14 articles report on this), and in these cases, gestures had a more pronounced manipulative focus. Where gestures were used as a stand-alone interface modality (58 articles) they conveyed both communicative and manipulative purposes, and it was difficult to classify them as only manipulative or only communicative. Thus while further sub-classification of manipulative gestures had not been encountered in literature, in order to explore the nature of gestures used in interactive interfaces, they were classified based on their contextual subcategory. This contextual subcategory definition is based on communicative gesture classification presented in Section 2. In this section gestures were classified based on the motions performed only, not considering the role they have in the application. Further analysis focusing on the role gestures play in the applications is given in Section 6.

Only six of the contextual sub-classes defined in Section 3 were identified in the reviewed articles: **semaphoric, manipulative, free-form, modalizing symbolic gestures, deictic, and pantomimic gestures**. Ten articles did not report on gestures used in sufficient detail, and for those it was not possible to determine the class of gestures. In some articles (Bourdot et al., 2010; Alves et al., 2015) it was clear speech was used, but not what it was used for specifically e.g. what was uttered and what the effect of that was. In the remainder of the articles where speech was used along with the gestures, it was not used in ways in which it would be expected based on the communicative gestures classification. It did not seem to semantically complement gestures so that they could be classified as iconic, metaphoric, or modalizing symbolic gestures. Speech instead tended to replace what could otherwise be achieved by semaphoric gestures, seemingly in order to simplify the implementation process. For example, when objects were rescaled word “scale” was uttered and then hands were used to indicate the amount of scaling. This will be further discussed in Section 6.

4.3. Prescribed and free-form gestures (instruction/definition based view)

The majority of articles reported on prescribed gestures (132 articles, 89%). This meant that the gestures were defined, and the participants in the study or the interface users were then asked to perform those exact gestures in order to interact with the system. A number of home appliance control interfaces (Liu et al., 2009; Zaii et al., 2015; Wu et al., 2016) and one 3D CAD application (Kela et al., 2006) initially allowed users to define their own gestures for particular activities, and then the most intuitive and most frequently used ones became the prescribed gestures. Similarly, Kim et al. (2008) used EMG to track four predefined dynamic gestures for robot control, selected from 20 initially tested gestures, based on ease of performance and quality of EMG signal they produced. This approach was a step forward as it considered

preferences of the users, which could lead to better acceptance of prescribed gestures. For anyone using the system however, other than the initial group of participants, the gestures used would still be prescribed. Future users of a system that the interface was developed for would not have been involved in the gesture elicitation, and would still have to learn the prescribed gestures. For the purposes of this article, we have classified those gestures as conditionally free. Six of the reviewed articles (4%) employed conditionally free gestures.

Free-form gestures were typically used to control an object, and they were encountered in 20 articles. For example, to direct a robot, create motion paths or surfaces, or modify virtual 3D object shapes or sculpt virtual clay. Free-form gestures were reproduced or copied into a system, and had a limited area of application. Symbolic activities, or activities that could not have been directly controlled, still needed to be facilitated using a different modality.

Again, a number of articles did not specify the gestures used. Where authors did specify if the gestures were free or prescribed the authors' classification was accepted, but in four articles there was not enough information on gestures to classify them. In Fig. 4 each box represents one gesture prototype, and the colour of the box indicated the type of gestures used. If an article reported on more than one prototype, each prototype was assigned a box. Prescribed gestures are placed above the horizontal axis, and conditionally free, free-form gestures and undefined gestures are represented below the horizontal axis.

It could be argued that fully free-form gestures would have the ability to convey a symbolic meaning in addition to the literal, copied shape, and would not have to be learnt prior to using a system.

5. Types of application, technology used, and gesture capture, tracking and recognition

How hand gestures are used in a specific interface appears to be influenced, to a certain level, by the type of the application and the purpose it serves, type of technology facilitating its implementation, and the underlying type of gesture capture, tracking and recognition supporting it. All three categories can be the inspiration behind the interface development that includes hand gestures. For example, gesture based interfaces were developed because they enable touchless interaction with medical imagery, reducing likelihood of contamination, or because they allowed for adjusting comfort functions in a car without taking eyes off the road, as an input method for workers using dirty hands, or who need to input information while holding tools in their hands etc. On the other hand, gesture interfaces were at times developed to demonstrate the capabilities of Kinect or LEAP sensors, or to demonstrate a new recognition method. Nevertheless, the nature of gestures used and manner of their implementation in an interface may uncover useful information for the research centred on gesture implementation. The arrangements of these three categories for different interface prototypes were highly varied. Within each a number of sub-categories were identified, and these and trends discovered within them

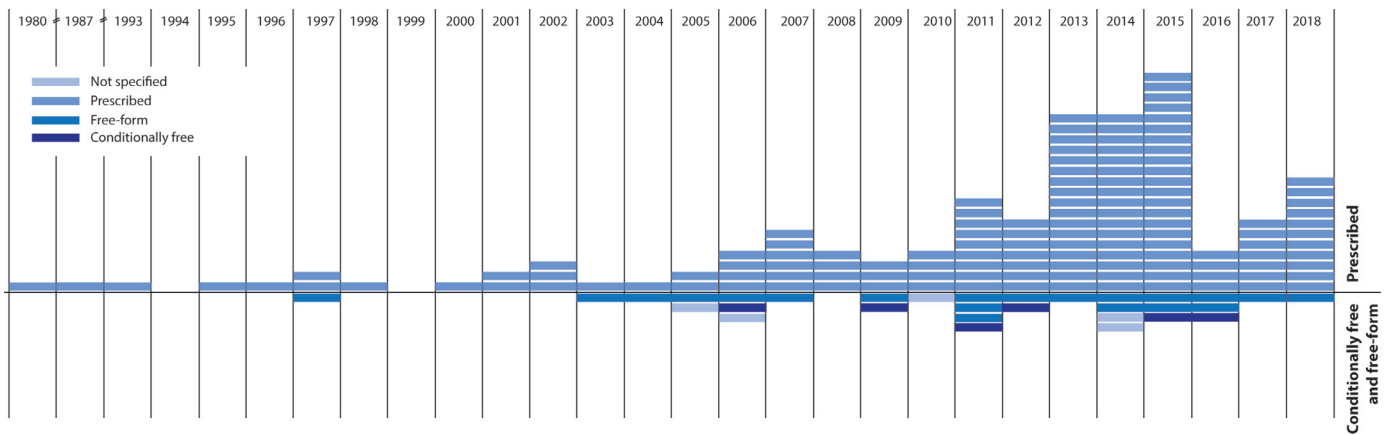


Fig. 4. Prescribed and conditionally free gestures.

are presented in this section. Then a more integrated approach is taken in Section 6, where they are viewed in conjunction with contextual type of gestures used in them.

5.1. Types of application

The first instance of gesture use in interfaces appeared in literature in 1980, envisaging a system for interaction with a display (Bolt, 1980). Then a prototype of a glove-based interface for 3D/VR object manipulation was developed in 1987. From 1993 to 2005 development of gesture-based interfaces was more constant, with one to three papers identified per year reporting on prototypes designed for various uses e.g. robot control, Computer Aided Design (CAD), manipulation of 3D objects, navigation/selection in an application, or developed interfaces for gesture recognition with no specific application determined. The significant increase of interest in gesture-based interfaces is noticeable in 2005/6, and has been further rising since 2013, likely due to introduction of a variety of depth-based cameras (see Section 5.2). While applications such as robot control, home appliance control, CAD manipulation, navigation/selection in an app, and game control were prominent, the fields of application were diverse. Different applications, ordered chronologically and arranged in five key categories are

shown in Fig. 5. Each box represents use of a specific application in one study, and majority of articles report on a single application prototype. It should be noted that this convention (one gesture based interface application, one box) was also used in Fig. 6. However, when the same set of gestures was tested using multiple technologies or for different applications, each technology or application was assigned a box. Numbers in the boxes indicate the type of specific technology used, listed in the legend.

The five categories of gesture-based applications have been identified and they were **3D modelling**, **assistive application**, **data input/authentication**, **manipulation/navigation**, and **touchless control**. The applications were categorised based on the immediate application, rather than the overall field of application. For example, a system could be used to provide a touchless gesture interaction through an interface providing interaction with 3D images. It would allow a surgeon to consult 3D scans during a surgery, and avoid the contamination of their hands. This application was classified as interface for manipulation/navigation as they were its immediate purpose e.g. the requirements gestures and the interface were developed to address predominantly focused on interaction with and manipulation of a 3D visualisation. Touchless interaction was a beneficial side effect, but not as influential as touchlessness was in an application providing remote robot control,

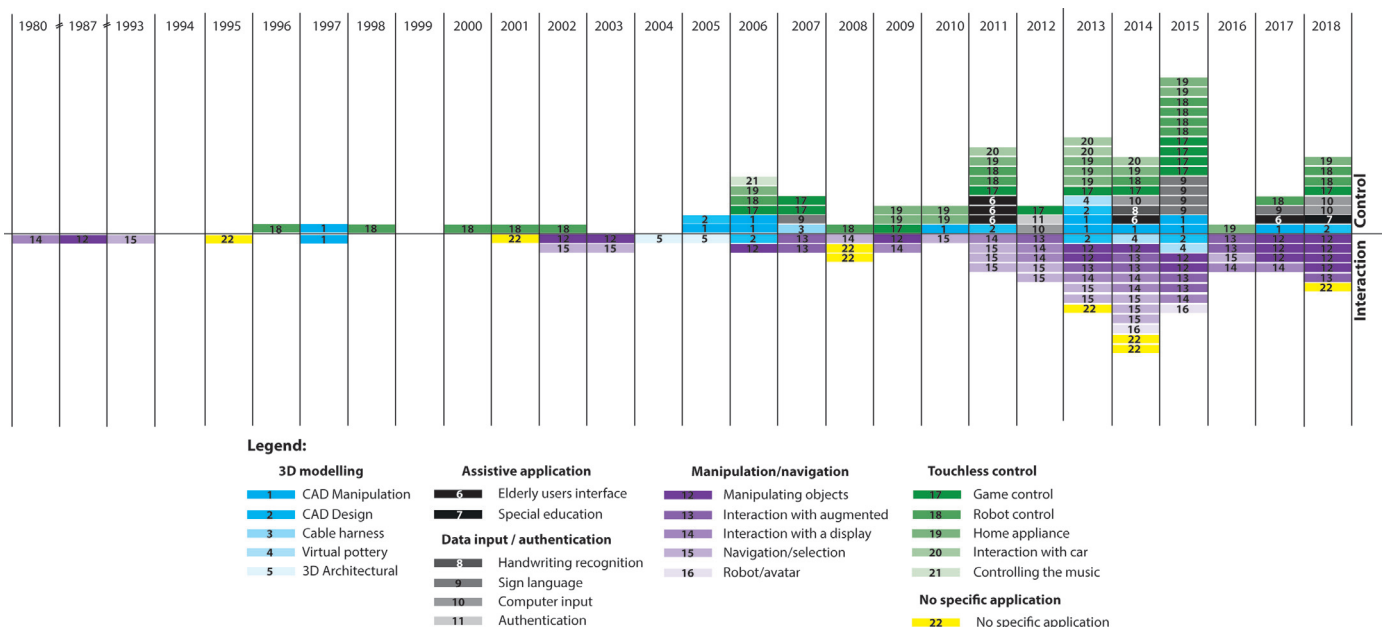


Fig. 5. Types of application for prototypes using gestures.

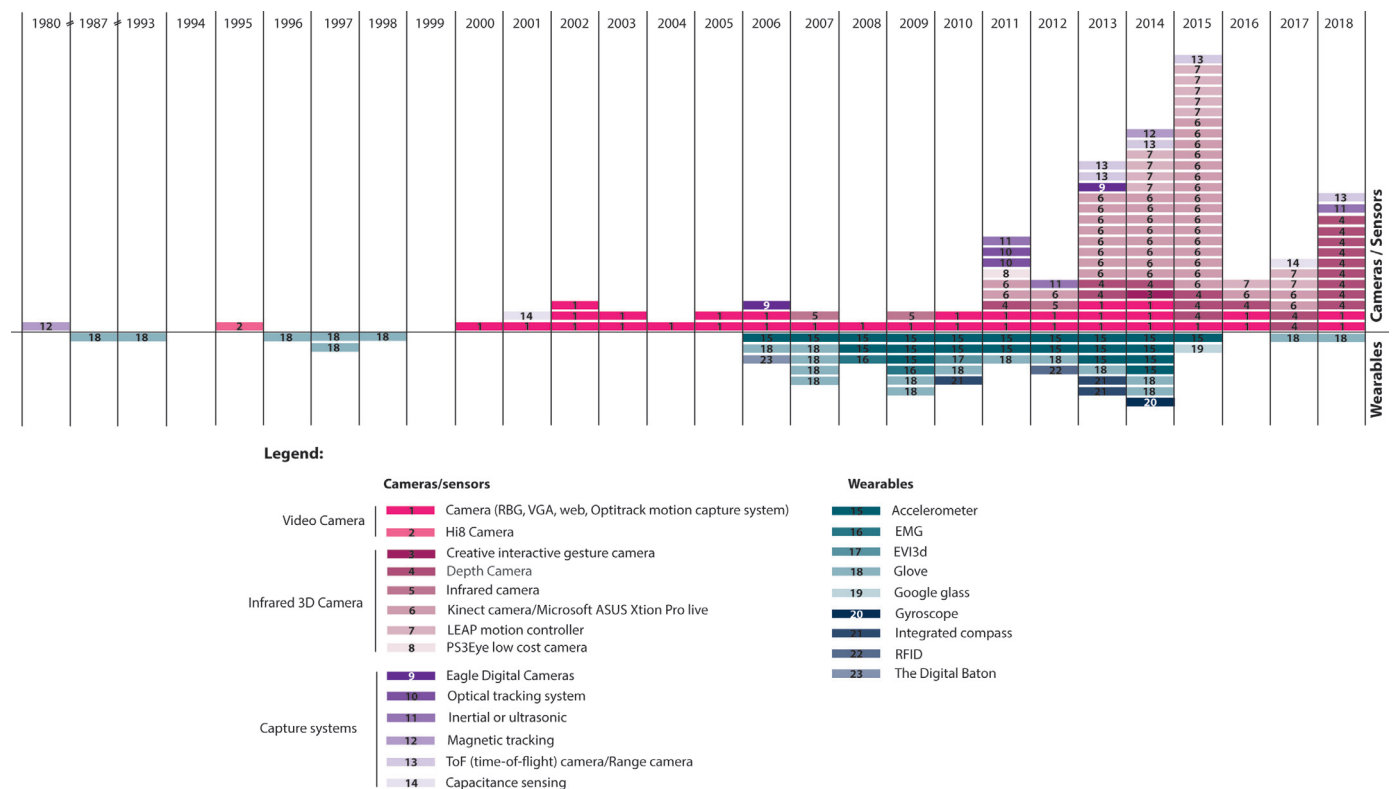


Fig. 6. Technology used to implement gesture interaction prototypes.

where robot was physically in a different location than the person controlling it. Overall application field was a medical field, but that did not influence the characteristics of gestures as significantly as the requirements for 3D interaction and manipulation did.

5.1.1. 3D modelling

Vinayak et al. (2013) provided a definition of 3D modelling activities, shape creation, modification and manipulation. They stated that creation is use of hand gestures to create a new shape in an empty working space, modification is the interaction with that shape intending to change its geometric characteristics, and manipulation is an activity that does not modify the shape itself but shifts it in space, translates, rotates or scales it. In this article, we refer to CAD design to indicate that shape creation and modification are taking place. CAD manipulation includes only manipulation of shapes, not their creation and modification. 3D modelling incorporates both, and includes all applications that have explored 3D design and manipulation of objects in 3D space, intending to develop a superior way of creation and interaction with these shapes. Use of gestures for 3D modelling is shown in Fig. 5, using boxes numbered 1–5. 3D modelling includes:

- 3D architectural urban planning (Buchmann et al., 2004; Yuan, 2005),
- Cable harness design (Robinson et al., 2007),
- CAD Design (Dani and Gadh, 1997; Kim et al., 2005; Qin et al., 2006; Holz and Wilson, 2011; Kang et al., 2013; Vinayak et al., 2013; Arroyave-Tobón et al., 2015; Huang et al., 2018),
- CAD Manipulation (Chu et al., 1997; Kela et al., 2006; Qin et al., 2006; Bourdot et al., 2010; Kang et al., 2013; Vinayak et al., 2013; Song et al., 2014; Beattie et al., 2015; Noor and Aras, 2015; Xiao and Peng, 2017), and
- Virtual Pottery (Dave et al., 2013; Han and Han, 2014; Vinayak and Ramani, 2015).

All of these applications were used to create, modify and manipulate

3D models, and were characterised by the dependence on 3D visualisation and the perception of 3D space. They often included interaction with the augmented or virtual world, but that was not their primary purpose. Majority used free-form gestures for the creation of splines or surfaces for the creation of the 3D model (Chu et al., 1997; Buchmann et al., 2004; Kim et al., 2005; Robinson et al., 2007; Holz and Wilson, 2011; Vinayak et al., 2013; Han and Han, 2014; Arroyave-Tobón et al., 2015; Vinayak and Ramani, 2015). Prescribed gestures were used to trigger predefined activities e.g. pinch gesture used to select an object or a grab gesture used to select an object, and prescribed gestures used in these applications were simple and emulated interaction with physical objects.

5.1.2. Assistive application

Assistive applications are marked with the number 6 in Fig. 5. These applications employed prescribed gestures to simplify the way elderly users interacted with electronic devices, computers, or robots providing aid in assistive living environments (Bhuiyan and Picking, 2011; Nazemi et al., 2011; Zhu and Sheng, 2011; Rodrigues et al., 2014; Carreira et al., 2017). All articles reported on early applications that showed promise, but there were no clear patterns in gesture use in them, as they mostly utilized a wide variety of predefined semaphoric gestures. One application explores the use of gestures in rehabilitation and learning within special education (Ojeda-Castelo et al., 2018).

5.1.3. Data input/authentication

Data input/authentication applications, numbered 7–10 in Fig. 5 were used for:

- Authentication (Guerra-casanova et al., 2012),
- Computer input (Cha and Maier, 2012; Yamada et al., 2014; Zeng et al., 2018),
- Handwriting recognition (Ammal et al., 2014), and
- Sign language input (Adamo-Villani et al., 2007; Kapuscinski et al., 2015; Liu et al., 2015; Santos et al., 2015; Trigueiros et al., 2015;

O'Connor et al., 2017).

Gestures were used to input information into a computer system, by using dedicated prescribed gestures, defined by interface designers or users, which have been repeated. For handwriting and sign language recognition, already established gestures were digitised, but they were also intrinsically prescribed.

5.1.4. Manipulation/navigation

Applications exploring more intuitive interaction, manipulation and navigation within different representation types are numbered with 11–15 in Fig. 5. They include:

- Interaction with a display/projection (Bolt, 1980; Choi et al., 2007; Foehrenbach et al., 2009; Beyer and Meier, 2011; Asadzadeh et al., 2012; Cauchard et al., 2012; Xie and Xu, 2013; Rossol et al., 2014; Saxen et al., 2014; Adeen et al., 2015; Wang et al., 2016; Braun et al., 2017; Osti et al., 2017; Dondi et al., 2018; Ma et al., 2018).
- Interaction with augmented reality (Reifinger et al., 2007; Lu et al., 2012; Bai et al., 2013; Hürst and van Wezel, 2013; Gangman and Yen, 2014; Adhikarla et al., 2015; Hernoux and Christmann, 2015; Shim et al., 2016; Saxen et al., 2014; Kim and Lee, 2016; Memo and Zanuttigh, 2018). Interaction with augmented reality included a variety of technologies that enable superimposed 3D representation of content and interaction with it.
- Manipulating objects in VR/3D (Zimmerman et al., 1987; O'Hagan et al., 2002; New et al., 2003; Deller et al., 2006; Moustakas et al., 2009; Wright et al., 2011; Djukic et al., 2013; Jacob and Wachs, 2014; Kim and Park, 2014; Al-Sayegh and Makatsoris, 2015; Covarrubias et al., 2015; Lopes et al., 2017; NICOLA et al., 2017; Park and Lee, 2018; Togootogtokh et al., 2018; Vosinakis and Koutsabasis, 2018). These were systems that primarily focus on manipulation of objects in either 3D or VR specifically, rather than unique AR solutions, or applications specifically aimed at CAD. Usually application is in medical imagery or VR facilitated interaction.
- Navigation/Selection in an application (Baudel and Beaudouin-Lafon, 1993; Krum et al., 2002; Wilson and Oliver, 2003; Wachs et al., 2008; Pang et al., 2010; Ni et al., 2011; Reale et al., 2011; Lin et al., 2012; Ruppert et al., 2012; Colaço et al., 2013; Riduwan et al., 2013; Fuhrmann and Kaiser, 2014; Widmer et al., 2014; Lee et al., 2016).
- Robot/avatar interaction (Alvarez-Santos et al., 2014; Alves et al., 2015).

Whether gestures were used to navigate 2D screens or applications, or interact with AR, VR or 3D space, one common trait of these interfaces was that nearly all gestures used were prescribed and predefined; they were defined either by interface designers, or in a small number of cases users initially could suggest preferred gestures for certain actions. Free-form gestures were used only for navigation through the space that was interacted with e.g. moving the mouse cursor, or moving an object that was virtually picked up. Additionally, the majority of the interfaces were multimodal, which meant that gestures had a limited breadth of application, as other modalities were used where appropriate.

5.1.5. Touchless control

Gestures used for touchless control of applications or objects are numbered with 16–20 in Fig. 5. They were used to control entities without physical interaction, allowing remote control and alternative, potentially safer modes of interaction. These applications include:

- Controlling a music recording (Lee et al., 2006),
- Game control (Carbini et al., 2006; Bannach et al., 2007; Kratz et al., 2007; Xu et al., 2009; Chen et al., 2011; Roccetti et al., 2012; Sodhi et al., 2013; Dardas et al., 2014; Lv et al., 2015; Santos et al., 2015;

Trigueiros et al., 2015; Yeo et al., 2015),

- Home appliance control (Kela et al., 2006; Liu et al., 2009; Schreiber et al., 2009; Chen et al., 2010; Pan et al., 2010; Boulabiar et al., 2011; Garzotto and Valoriani, 2013; Hoste and Signer, 2013; Takahashi et al., 2013; Dinh et al., 2014; Denkowski et al., 2015; Zaii et al., 2015; Wu et al., 2016),
- Interaction with car controls (Mahr et al., 2011; Kajastila and Lokki, 2013; Riener et al., 2013; Lauber et al., 2014; Buddhikot et al., 2018),
- Robot control (Pook and Ballard, 1996; Savage-Carmona et al., 1998; Waldherr et al., 2000; Fong et al., 2001; Rogalla et al., 2002; Hasanuzzaman et al., 2006; Kim et al., 2008; Van Den Bergh et al., 2011; Xian et al., 2012; Gil et al., 2014; Boboc et al., 2015; Cicirelli et al., 2015; Marasovic et al., 2015; Xu et al., 2015; Kim et al., 2017).

Gestures used in these applications were, similarly to those used for interaction with different representation types, a mix of predefined gestures triggering predefined actions, which have to be learned, and free-form gestures for navigation between two predefined gestures. Free-form gestures were typically used more for robot or game control. Prescribed gestures occasionally took into account user preferences during gesture elicitation, usually for the home appliance control gesture definition.

5.1.6. No specific application

Occasionally gesture interfaces were tested in a prototype, but specific application was not determined, and these are designated with number 21 in Fig. 5. These applications all used prescribed gestures defined by the interface developers, and focused on technology facilitating the interfaces and recognition technology more than the gesture elicitation (Quek, 1995; Rekimoto, 2001; He et al., 2008; Niezen and Hancke, 2008; Palacios et al., 2013; Huang et al., 2014; Zhou et al., 2014; Kim et al., 2015; Liu et al., 2018). The focus was on testing different combination of gestures and technologies in order to determine the potential for their use in the future.

5.2. Technology used

Technologies facilitating gesture-based interfaces can be split into two overall groupings: those relying primarily on a camera or sensor based tracking, where users hand gestures are recorded at a distance, and those relying primarily on wearables, where users wear devices such as gloves, rings, bracelets, bands with accelerometers etc. The former can also be referred to as visual-based, and the latter as physical based. The key difference between them is that visual based do not require the user to wear any kind of technology and they perform gestures using bare hands, whereas wearables require use of physical devices mounted or worn on the hands. In Fig. 6 the camera/sensor facilitated solutions are represented above the horizontal line, and wearables-facilitated solutions below the line. Each box represents use of a specific technology in one study, and numbers in them serve to indicate the type of specific technology used, listed in the legend.

5.2.1. Visual based technologies

Within the camera and sensor grouping, there were three distinct sub-groups of technologies:

- Regular video cameras,
- Infrared/depth 3D cameras,
- Motion capture systems:
 - 3D optical motion capture systems,
 - Laser based tracking,
 - Magnetic based tracking, and
 - Capacitance sensing.

Technologies were referred to by different names in different studies, and that terminology was retained if they were discussed in detail in this review, but they were grouped according to the underlying technology that enables them, as shown in the legend in Fig. 6. For example, creative interactive gesture camera, depth camera, infrared camera, Kinect camera, Microsoft ASUS Xtion, LEAP sensor and PS3Eye are based on infrared waves enabling depth perception, so they were classified as Infrared/depth 3D cameras. While the majority of technologies in this category did not require anything to be worn to aid detection, video cameras often necessitated use of markers placed on hands and fingers. It is noticeable that the introduction of infrared cameras, particularly Kinect cameras and LEAP sensors, had inspired a large number of gestural interface studies in the recent years, as 57 out of 148 studies (39%) use Kinect, LEAP or ASUS Xtion. 65 out of 148 studies (44%), used some type of infrared camera. Kinect and LEAP have the benefit of being easily transportable and the simplicity of their use, use of bare hands, and lack of interference with the users' gestures aligns well with the goal to achieve intuitive interface interaction. The drawback of this technology is that, while it has improved significantly in the recent years, it still does not possess the consistent precision and reliability required to provide sufficiently accurate gesture tracking, capture and analysis. For example LEAP has been found not to be sufficiently robust for use in medical clinical studies (Coton et al., 2016). Cameras and sensors also suffered from occlusion and configuration complexity problems (Rautaray and Agrawal, 2015). While the functionalities of the cameras, sensors and motion capture systems listed above were based on different principles, and some provided higher accuracy, their capabilities in terms of gesture recognition were similar.

5.2.2. Wearables

Wearables included gloves, accelerometers, markers, radio-frequency identification (RFID) based devices, compasses, gyroscopes, electromyogram (EMG) based devices, google glasses, or bespoke solutions for gesture recognition such as EVI3d or The Digital Baton. These were either specific products that a user can wear, or which required the user to hold a device in their hands. It has been reported that wearing or holding a device while performing gestures can influence the users, as it can physically limit them in different ways. It could also lead to a higher degree of fatigue compared to using technology that does not require physical wearables. Additionally, wearables could simply be uncomfortable for the user (Rautaray and Agrawal, 2015), which was an influence in itself as it could detract users' attention away from the gestures. Since the gestures were often not used as a sole interaction method, but a part of a multimodal solution, the wearables could introduce further restrictions though propagating their limitations to other modalities. On the other hand, wearables often provided higher accuracy and suffered from less configuration complexity, compared to Kinect and LEAP for example. The most frequently used technology, within the wearables category, was the accelerometer, used in 22 of the studies, as a main technology in 17, or supporting technology (along with cameras, integrated compass or EMG), in the remaining ones.

5.2.3. Multimodal approach

70 articles reported on multimodal approach, i.e. employed two or more different technologies to achieve desirable gesture tracking, recognition and integration in an interface. In some cases, this was due to current technology not being advanced enough to achieve the desired accuracy and speed of tracking. In other cases, the reasoning behind the use of additional technology was to better support different modalities of communication and interaction that complemented gestures, and provide a more suitable solution for the interface being developed. While supporting technologies were varied, speech, Head Mounted Display (HMD) and VR seemed to be the most frequent ones. For 3D modelling applications there seemed to be a slight preference for camera/sensor solutions, particularly infrared cameras. If wearables

were used, glove was the most frequently used technology, particularly supported by the use of VR. Preference towards sensors and cameras was likely supported by the freedom of movement and the ability to perform gestures in a manner that was uninhibited by the technology used. Gloves on the other hand provided higher accuracy. Their benefits and drawbacks seem to have influenced the choice of technology use significantly. Suitable visualisation was an important aspect of 3D modelling, as VR or AR allowed 3D gestures to be matched with 3D representation. Interaction and manipulation with different representation types were more commonly performed using video and depth cameras, very often supported by speech interfaces or visualisation aids such as VR, HMD, or projections. For the remaining applications, no specific trend was discernible.

5.3. Gesture capture, tracking and recognition

There are three fundamental phases required to translate a gesture from a physical activity into an input for an interface: detection, tracking and recognition (Rautaray and Agrawal, 2015). Frameworks for detection and tracking seemed to be highly dependent on the capabilities of the technology used. In the reviewed articles, recognition was performed using a variety of established algorithms and original algorithms different research groups were developing for their purposes, supported by the use of established libraries or SDKs (software development kits). Some of the algorithms used were Active Appearance Model (AAM), Distance Metric Learning, Dynamic Time Warping (DTW), Final State Machine (FSM), Hybrid approach for gesture recognition with depth maps (HAGR-D), Protraction, Rule Induction, Support Vector Machine (SVM), Uni-stroke, and Spatiotemporal Convolution Neural Network (ST-CNN). One of the most common recognition approaches was Hidden Markov Method (HMM) (Quek, 2004), and 17 articles (11%) used it for gesture recognition. Original algorithms were reported on in 31 articles (21%). These were algorithms developed specifically for the needs of the solution that was being reported on. It is possible that incidence of original algorithms was even higher, as a number of articles mentioned use of programming languages, but did not specify the recognition technique. However, this assumption has not been made, and when not clearly stated what technique was used those articles were classified as not specified. 38 (25%) articles were using SDKs, and a number of different ones were used, such as Android SDK, Natural Interaction Middleware (NiTE), Open CV, Open NI, OPIRA, Point Cloud Library (PCL), and Unity. While there are differences between them, they are all widely adopted and standardised recognition frameworks. They were used along with an algorithm, and there seemed to be no specific correlation between the SDK used and the algorithm, even when the applications were very similar. Some applications used software specifically developed for the technology used, which performed the recognition, such as EVaRT 4.4 software (Qin et al., 2006) or EVI3d software (Bourdout et al., 2010). It appeared that in the majority of cases the recognition technology was a pragmatic choice of technology/software that allowed capture and provided the necessary accuracy. Therefore, the recognition method used was not considered in further analysis. Due to technological developments, some of the recognition methods mentioned in this section are now largely outdated e.g. the need for NiTE/OpenNI depth sensors based recognition is declining. Therefore it is interesting to note that in the past two years the most used recognition methods tended to focus on original algorithms (often written in Python), SVM, Neural Networks (NN) and use of Unity.

6. Gesture interpretations

In Section 4.2 gestures were classified contextually based on the motion performed by the hands. In this section the role gestures have in an application is considered and how it affects the contextual classification. A Venn diagram was used for the analysis, shown in Fig. 7, and

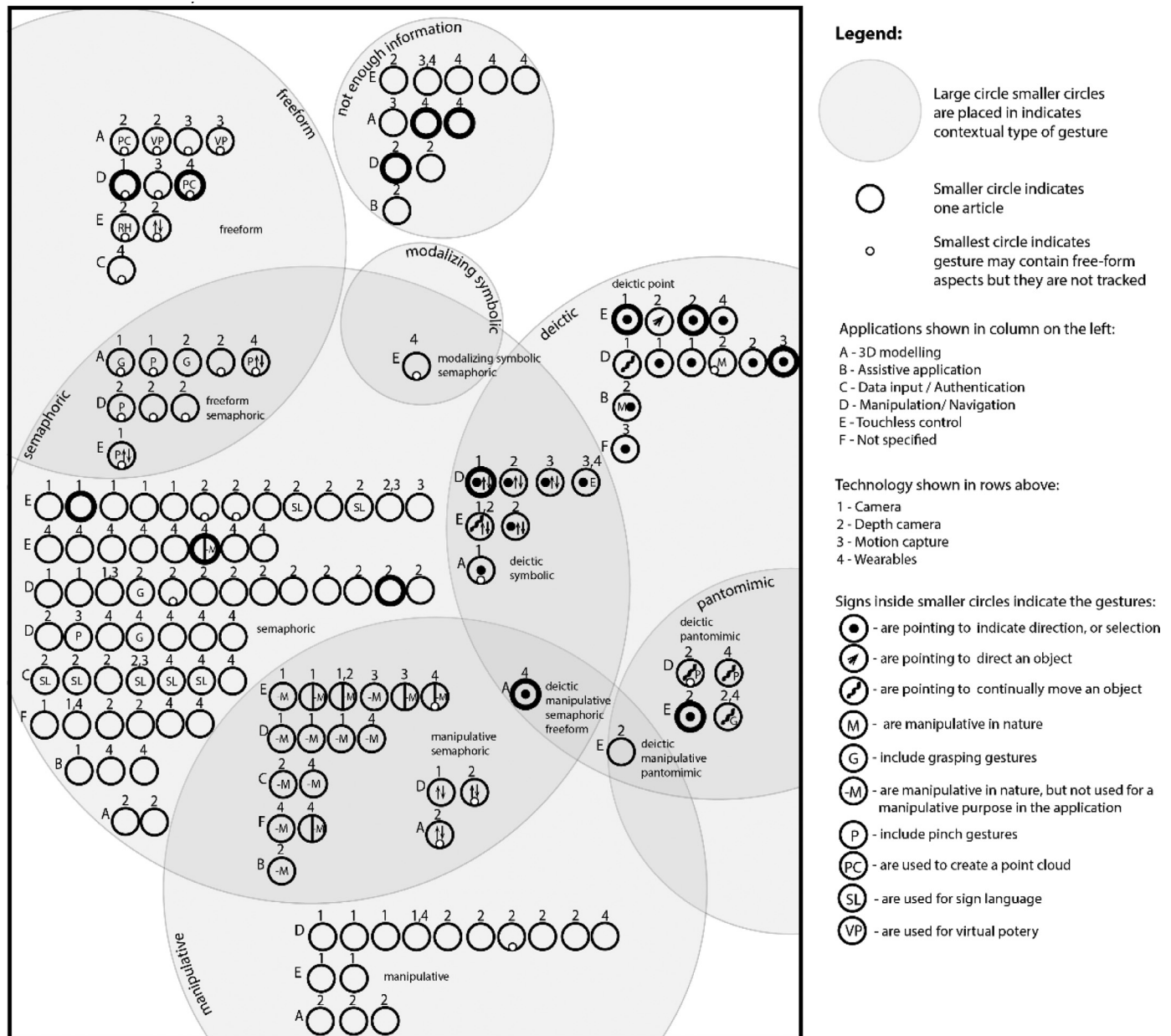


Fig. 7. Venn diagram nature of gestures, their role in the application and the technology used.

based on a type of gesture used in an application, the article reporting on it (represented by a small circle) was placed in one of the large circles representing one of the six contextual gesture types identified in Section 4.2. Then both types of circles were arranged so that articles implementing the same gesture types were grouped together. This has identified 11 groups of gesture types. The darker the shade a specific zone is; more types of gestures were used in it.

Some applications report on exclusive use of deictic, manipulative, free-form, or semaphoric gestures. However, many of the applications combine more than one type of gestures. In total 11 distinct gestures or combinations of gestures were identified:

- Deictic
- Deictic and pantomimic
- Deictic and semaphoric
- Deictic, manipulative and pantomimic
- Deictic, manipulative, semaphoric and free-form
- Manipulative

- Manipulative and semaphoric
- Free-form
- Free-form and semaphoric
- Semaphoric
- Semaphoric and beat.

Within each region gestures were also arranged based on the application they were used for (denoted with letters A to F to the left of the row of circles, as listed in the legend) and based on the technology facilitating them (denoted with numbers 1 to 4 above each of the circles). Where speech was used as one of the modalities, small circles have a thick border. Only 14 applications used speech.

It is noticeable that semaphoric and deictic gestures are present in most of the large circles in the diagram. Deictic gestures often provided positioning and selection function, while semaphoric gestures added predefined, often abstract added functionality that could, but did not have to correspond in meaning to the gesture performed. Looking at the relationship of speech and gestures, four types of relationships are

identified:

- 1 Point and command approach, where pointing gesture is used to indicate which object or direction is referred to, and then a host of predefined utterances are used as a command to interact with the object (Bolt, 1980; Carbin et al., 2006; Hoste and Signer, 2013; Choi et al., 2007; Chu et al., 1997; Wilson and Oliver, 2003; Moustakas et al., 2009; Pan et al., 2010; Dani and Gadh, 1997; Bourdot et al., 2010; Boboc et al., 2015) e.g., user can point at an object, utter “Move that there” and point the second location, in order to move an object. Or use commands like ‘select’, ‘dimension’, ‘scroll’, ‘detach’, ‘attach’ on it, and commands like ‘ok’ or ‘release’ to cancel or complete the activity. This approach is consistent with the role deictic gestures play in speech and gesture theory, but it is constrained to the use of predefined commands.
- 2 Replacement of gesture approach (Rogalla et al., 2002; Dani and Gadh, 1997; Bourdot et al., 2010) – where speech is used to completely replace the gesture, not complement it. For example instructing a robot to “take the cup from the table which is in front of you”, or command ‘create block, length 10 width 10 height 5’ to create a block or provide verbal coordinates it should be moved to. (Bourdot et al., 2010) developed an ontology processor is used as interface between speech recognition system and a targeted application, which allows use of natural speech which is then transformed to commands. These speeches replacing the gestures are not necessarily the most fitting ones, they seem to focus on convenience rather than the best fit for an activity.
- 3 Text input approach (Choi et al., 2007) – where words uttered are automatically recognized by the computer, and gestures are used to correct any errors or move words around to achieve the content desired. In this approach speech and gestures are completely separated, as speech is only used for text input. Gestures also do not have a communicative or manipulative purpose.

In ten articles information on the specific gestures performed was not given, making it impossible to classify these articles. In two of these articles speech was used (Alves et al., 2015; Fuhrmann and Kaiser, 2014), and it was not specified what role the speech played either.

Regardless of the speech link, in the following sections gesture classes will be analysed based on the way they are used in an application.

6.1. Deictic gestures and their combinations

Deictic gestures were used in one of two ways, a pointing gesture was used to indicate direction or the selection of a point, or a pointing gesture was used to continually move an object along a path created by that hand. One finger, two fingers or a fist were used as a pointing device. In Fig. 7 pointing in order to select is marked with a dot in the middle of a circle, pointing in order to direct is marked with an arrow in the middle of a circle, and pointing in order to move an object along a path is marked with an M in the middle of a circle. A combination of **deictic and pantomimic gestures**, was used in a number of applications for interaction and manipulation of different representation types or for touchless control. A deictic gesture was used for pointing and selection in one application, and in the remaining three it was used to move a cursor or an object. Pantomimic gestures used were swiping, pinching and grasping gestures used to pick up and pull and modify parts of objects. They were classified as pantomimic as they emulated the real life activity that would have been applied to a similar object. A combination of **deictic and semaphoric gestures** was used for the same purpose, but deictic gestures were pointing gestures used for selection, and different predefined semaphoric gestures were used for manipulation or triggering specific commands. A combination of **deictic, manipulative and pantomimic gestures** was used for a touchless control of a game character (Sodhi et al., 2013), where deictic

gesture was used to indicate where the character is, swiping, manipulative gesture to intercept the virtual button, and an pantomimic pushing gesture was used to push a virtual button. A combination of **deictic, semaphoric, manipulative and free-form gestures** were used for an early 3D manipulation application (Chu et al., 1997), where pointing was used to select, grasping to move/zoom, a number of semaphoric gestures to create 3D objects, free-form gestures to create surfaces and manipulative gestures to change the object dimensions.

6.2. Free-form gestures and their combinations

Free-form gestures were used as a sole type of gestures in two variants: as completely unrestricted gestures that are reproduced by the system tracking them, essentially copied in as a path or a surface, and as completely unrestricted gestures that were emulated by a virtual hand created in a software system. The majority of applications were 3D modelling and touchless control. Some 3D modelling applications worked on a principle of creation of a point cloud as the hands move, that was subsequently recognised as a shape. Some were based on virtual sculpting, and a virtual shape on a virtual turntable was being modified based on the positioning of the hands. For touchless control applications, free-form gestures were tracked and used to move virtual objects, windows or pointers, or to control a robot hand. Combination of **free-form and semaphoric gestures** was used for the same types of application, but the semaphoric gestures were added to allow for added functionality, specific to each application. However, overall they were used to trigger a predefined activity. In the combination of free-form and semaphoric gestures, free-form gestures used were completely unrestricted and mostly used for selection or positioning of objects. Gestures that contained elements of the free-form movement, but where those elements were not significant enough for gestures to be classified as free-form as they were not fully tracked, were denoted with a smaller circle added to the bottom centre of the circle standing for an application in Fig. 7.

There were no truly free gestures employed in the prototypes tested. Vinayak et al. (2013) created shapes by sweeping a profile in space along a path using hand gestures, modifying them using free-form and parametric deformation and manipulation. Kim et al. (2005), Qin et al. (2006), and Holz and Wilson (2011) used free-form gestures for surface creation, where the surface motion of the hand traces would become a new surface or spline. Dani and Gadh (1997) developed a twofold system, with a free-form surface design mode for parts that cannot be created using standard shapes, and parametric design mode used to create standard shapes by combining hand position and voice commands. While unrestricted in terms of the shapes and paths which can be created in these types of systems, overall free-form gestures used in interfaces developed so far still have other types of limitations. Any symbolic or metaphoric activity still needs to be achieved by either performing prescribed gestures or utilising a different modality of interaction, such as speech or pressing a button on an interactive device. One exception was the article by Kim et al. (2017) where robot was trained to perform an action that was predefined, but the gestures different users performed to train the robot for that action were free and could vary. The system then learned from each of the users for each of the predefined activities, and added the information to a database. This would lead to a system that is individualised for each person, but only for a number of predefined activities.

6.3. Manipulative gestures and their combinations

Manipulative gestures were in most cases used for touchless interaction, and were based on a set of gestures aimed at achieving translation, rotation, scaling/zooming, or object size manipulation. Their goal entirely corresponded to the motion performed by the hand. However, specific applications did define their own motions which needed to be repeated in order for the gesture to be recognized. These

were semaphoric in the sense that size gestures or paths travelled by the hand did not always directly correspond to the achieved modification upon the object controlled. For example to scale the size of an object the index finger and thumb were moved closer together, and the movement of a few millimetres could correspond to the size reduction an order of magnitude larger. Combination of **manipulative and semaphoric gestures** had the same application, but the semaphoric gestures provided additional predefined functionality that could be triggered by performing a specific gesture. Variety of other types of gestures had elements which were manipulative in nature, and in Fig. 7 these are denoted with arrows pointing up and down. In these cases the aim of the gesture is not a manipulative activity, although the paths followed by the hand resemble a manipulative gesture. This is further described in the following paragraph describing semaphoric gestures which are manipulative in nature.

6.4. Semaphoric gestures and their combinations

Semaphoric gestures were the most frequently used gestures, and they functioned on the basis of a user performing an abstract, predefined motion representing a concept using their hands, that triggered an assigned predefined activity. The gestures needed to be learned and performed accurately to be recognized. Semaphoric gestures were used in all of the application types identified in this review, and in 70 applications, they were used as a sole type of gesture. Gestures used differed from application to application, with the exception of those that used existing languages such as sign language or hand writing recognition, in which 30 letters of alphabet were used as symbols for recognition. A specific subsets of semaphoric gestures were those in which the **semaphoric gestures performed are manipulative in nature** but are used to trigger a predefined activity that was not manipulative in nature. That is, the nature of the activity was not linked to the nature of the symbol. For example, moving a hand upwards vertically in order to turn a TV on. In some of the applications in this subset, all of the gestures performed were manipulative in nature, but did not trigger manipulative activity; in others only a limited number of gestures had this nature.

A combination of **modalizing symbolic and semaphoric gestures** were used in only one application (Lee et al., 2006) to control a music recording, and the beat gesture was an auxiliary gesture. Gestures used were predominantly semaphoric. Two semaphoric gestures emulating a music conductor were used and recognised, with additional speed tracking which controlled the speed of music played in the application. A third gesture was not predefined, and only its speed/beat rhythm was tracked.

7. Patterns of gesture use

While there were some technologies that were used more prominently for some of the application types, a clear prescriptive link between the technologies used, gestures performed and the applications they were used for could not be made. However some patterns were identified in the ways in which use of specific gestures, applications and technologies are combined together, and they are shown in Fig. 8. They were derived by taking a higher level view of the Venn diagram given in Fig. 7. Within the 11 zones of identified gesture types, articles were first grouped based on the application they are used for, discarding the outliers (applications with one or two articles reporting on them, in a zone where other applications are reported on in a significantly higher number of articles). Then these application based groupings were linked across the 11 gesture type groupings based on the facilitating technologies used in majority of the articles in each of groupings.

Visual based technologies were a dominant type of technology used for gesture based interfaces. The use of gestures in interface expansion seems to have been inspired by the availability of affordable depth cameras and motion sensors, which might have had an influence on the

frequency of their use. Nonetheless, visual recognition of gestures was more present for manipulative, deictic, pantomimic gestures and their combinations. Free-form gesture based 3D modelling, and combination of free-form and semaphoric gestures for 3D modelling were also more often facilitated by the cameras, 3D cameras and motion sensors. When wearables were used for these types of gestures, free-form gestures in particular appeared to be most commonly paired with gloves. These choices were likely due to either freedom of movement allowed for by the cameras and sensors, or accuracy and ease of capture if glove was used. Where complex combinations of gestures were used and each had a specific role to play, accurate recognition became more important, and wearables were primarily used in these applications. In comparison, applications using semaphoric gestures exclusively tended to use wearables more frequently than applications using other types of gestures, but cameras and motion capture sensors were still more prominent. In these applications around 2/3 of supporting technologies were visual based, and 1/3 were wearables.

8. Potential ways forward for gesture characterization and elicitation

The goal of this review was to investigate the patterns of touchless hand gestures used in gesture-based interfaces, which have reached the prototype stage. In this process, the reasoning behind the elicitation of these gestures was explored. It appears that a framework guiding the gesture elicitation does not yet exist, not on the higher level, in terms of which aspects of interfaces would be best supported by gestures, nor for specific fields of application, in terms of how to select the most appropriate gestures for specific activities. During the mapping of the field of gesture interfaces observations were also made on gesture classification process and influence the technology has on which gestures are used. They will be discussed in this section.

8.1. No accepted framework

Research in the field of gesture based interfaces is extensive, specifically in fields of application such as touchless control and interaction with different implementation types. This is particularly true for home appliance control, robot control, game control and medical imagining manipulation. However, it seems like development of a framework or contribution to a standard for gesture use for specific applications is still not imminent, although there were developments moving in that direction. For example, Riener et al. (2013) collected 400 micro gestures for eight car functions explored and recorded using Kinect, with further evaluation pending for recommendations towards the standard. However, to date, no standards for gesture elicitation or application in gesture based interfaces for a specific field have been identified.

While discussions on applicability of gestures and appropriateness for the application were present in a number of articles they were limited, and did not seem to significantly influence the way gestures used were selected. One reason for this is certainly the diverse range of applications, as due to the differences in requirements the frameworks for gesture elicitation would likely take different forms. Nevertheless, general lack of reasoning behind the gesture selection might insinuate that gestures could perhaps be optimized in the future, for a specific application, or a group of similar applications. There are however, activities present in the review papers that are building towards standardisation or a framework, particularly in the articles that report on initially user generated gestures. While those articles report on work implementing gestures that were defined based on user preference, the sample sizes used to define them were small. Majority of articles, 46, included 10–20 participants. Only four articles had included more than 60 participants, 19 articles included between 20 and 40 participants, and 25 included less than ten participants. Among the articles that included larger number of participants there are those like (Guerra-

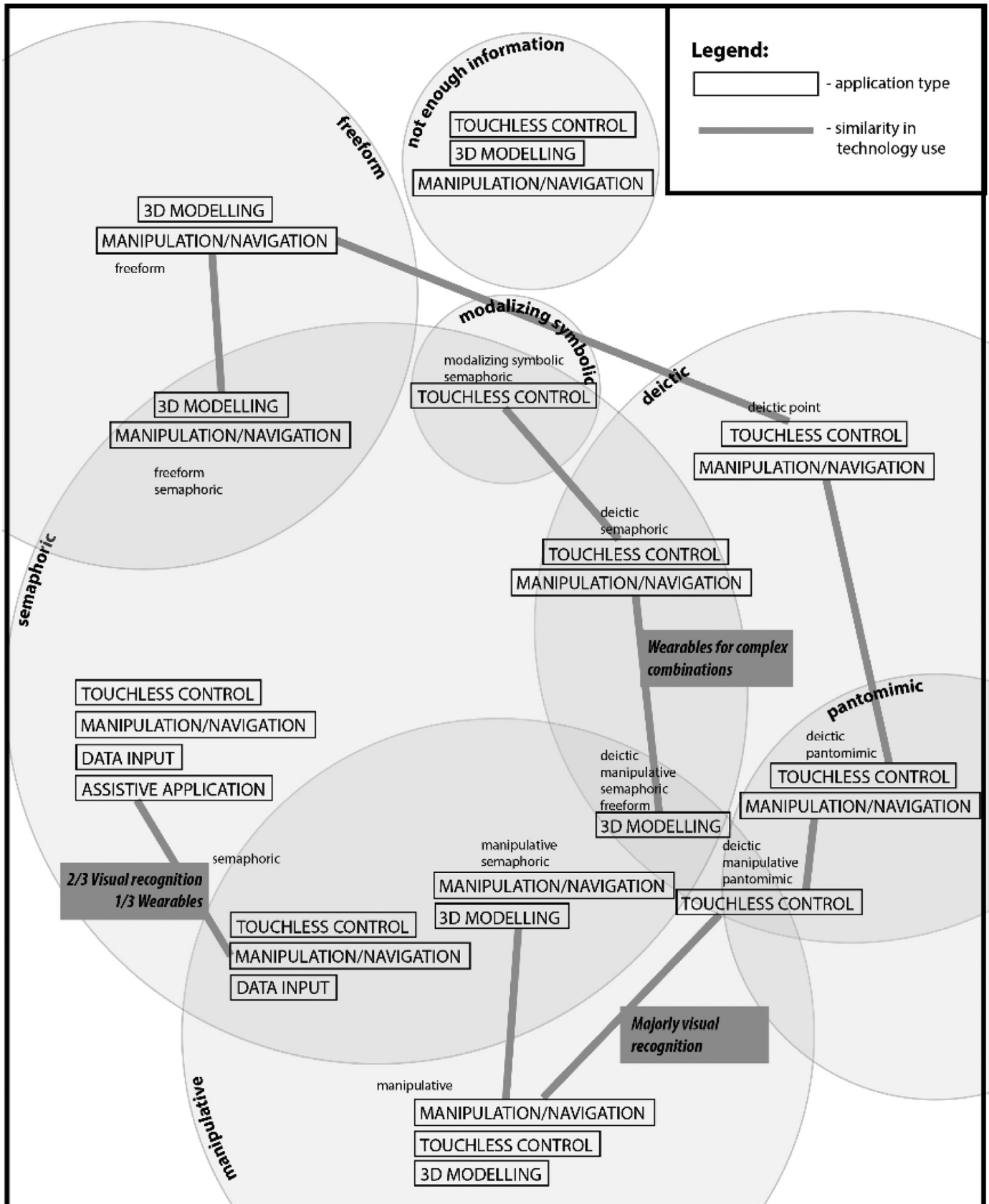


Fig. 8. Patterns in gesture use corresponding to technology and applications they were used with.

casanova et al., 2012) that used 100 participants that tested an authentication application where each participant used only one gesture, to unlock a phone, and the gesture was different for each of the participants. As the gesture had triggered a predefined activity, the important aspect was recognizing each individual gesture, not comparing them or building towards a framework. Kim et al. (2017) on the other hand allowed participants to use gestures freely, and the system was adding them to a learning database for three predefined activities. The application field was too narrow for the gestures to be generalised across a wider number of interfaces, but applications like this one have the potential to extract valuable information for future development of the gesture elicitation framework in its field. Within design research, discussions regarding what is considered statistically meaningful and how reliable and justifiable results can be achieved is already happening. There is a need for common standards in design science which could support key quantitative methods such as meta-analysis, that could lead to generalizable insights, applicable across the related research fields (Cash, 2018). Similarly, guidance on required number of participants, their background and skills, and how that affects the gesture elicitation process would benefit the field of gesture research and application.

In addition to low numbers of participants, experiments performed to test these applications often did not follow an in-depth evaluation process to determine the appropriateness of the gestures for the specific application, or their efficiency. This is not a criticism of the reviewed studies, as their primary goal was successful initial implementation. However, in order for interfaces to gain widespread adoption, gesture suitability is an important aspect that is currently not readily available in the literature. Development of a framework for gesture elicitation would help guide further gesture interface development, and ensure important aspects are identified and considered. Any framework would likely have to be specific to a context it is developed in, in order to provide suitable support to gesture elicitation and interface development based on it.

Frameworks that exist in the context of gesture research tend to focus on gestures closely linked to speech, or focus on technology and not gestures per se. For example, Cassell (1998) developed a framework for the generation and interpretation of spontaneous gestures in the context of speech, that is not pre-defined but creates natural interfaces. In it, gestures were defined as hand gestures that co-occur with spoken language. Interaction with very visual interfaces such as VR/AR and 3D object manipulation, or interaction with robots, avatars or home appliances in 3D world, while communicative in nature, does not always need to include speech, and requires existing frameworks to be extended to cover more than just speech linked gestures.

8.2. Low cross-pollination between similar solutions

Research groups publishing in the field often build on their own previous work e.g. Vinayak et al. (2013) and Vinayak and Ramani (2015) show a progression in development of a virtual pottery solution. Likewise, technology considerations in the field are referenced in the background searches, which articles report on. All reviewed articles have referenced work already present in the field, in terms of the technology used and key concepts introduced. However, none of the articles seemed to compare the gestures they have selected for use in an interface with gestures used by others performing similar research in the past. For example, Kela et al. (2006), Chen et al. (2010), Hoste and Signer (2013), Takahashi et al. (2013), Zaii et al. (2015), Wu et al. (2016), and Buddhikot et al. (2018) all explored use of gestures to control a TV. While Wu et al. (2016) referenced work performed by Hoste and Signer (2013), and compared the goals of the applications developed, there is no evidence of consideration of specific gestures used by earlier work in these articles. Work towards a framework for gesture elicitation could ensure important aspects identified by different groups during gesture selection are considered in the future

gesture interface development.

An article was identified through the review, by Jahani and Kavakli (2018) that is an example of good practice. It did build on previous work in the field, attempting to find patterns of fully free-form gestures users perform while describing two different chair designs. The authors built upon and further developed an existing coding system (originally by Wobbrock et al. (2009)), and identified patterns performed by hand gestures. These gestures were not implemented in an application; hence, the article did not conform to the inclusion criteria of this review.

While the interface types in the field are diverse and likely require different frameworks and approaches, similarities could likely be identified within related sub-groups. This could lead to a development of frameworks that, if they were general enough, could allow for further expansion in order to fit the needs of the specific applications being developed.

8.3. Contextual classification issues

Gestures were used as a sole mode of interaction or as one of the modes in the multimodal interface. However, it seems like their role did not change extensively depending on if they were sole or one of the modalities. In fact, it is possible that convenience linked to current capabilities of the available technologies was an influential argument behind the reasoning for the choice of the multimodal technology use. In multimodal solutions, speech and buttons were seen to supplement gestures. However, it seemed that speech did not supplement the gestures in the way it may be expected, based on the knowledge about the correlation between speech and gestures originating in communicative gesture based research. That is, speech and gestures were often used to perform the same activity, and which one was used was decided based on the ease of technical application, case per case, rather than based on the role speech and gestures are known to have. This may indicate that a different categorisation is needed for gestures that are used to express a communicative purpose that relies heavily on spatial characteristics and interaction, in addition to conveying metaphoric and symbolic meaning.

Observing the theoretical background on gestures, which is still used for classification in numerous articles reviewed, it is clear that the approach taken to classification originates in observing gestures and the role they serve during the verbal communicative process. Even the classification adopted in this review is based on classification originating in speech and research of gestures in the context of speech. However only 14 out of 148 applications analysed have actually used speech in their interfaces. While gestures in them do serve a communicative purpose, it is not necessarily related to verbal communication, and perhaps a development of an approach that decouples gestures from speech, and observes them on their own could be beneficial. This could contribute a different view with different results, that could potentially be integrated with the existing knowledge about gestures used along with speech. These could potentially contribute to the field of gesture research by providing findings based on gestures focusing on ergotic (gestures used to manipulate the physical (or virtual) world (Billinghurst, 2018) and epistemic (gestures used to learn from environment through tactile or haptic exploration (Billinghurst, 2018) gestures, enriching the gesture classification and uncovering types of gestures present in interfaces enabled by developing technology.

The findings that both approaches contribute could then be compared, and possibly form a new sub field of gesture research, as shown in Fig. 9.

8.4. Dependence on technology

A variety of different applications and technologies are used for gesture interface implementation, and each combination of approaches can rely on different types of gestures and roles they serve in an

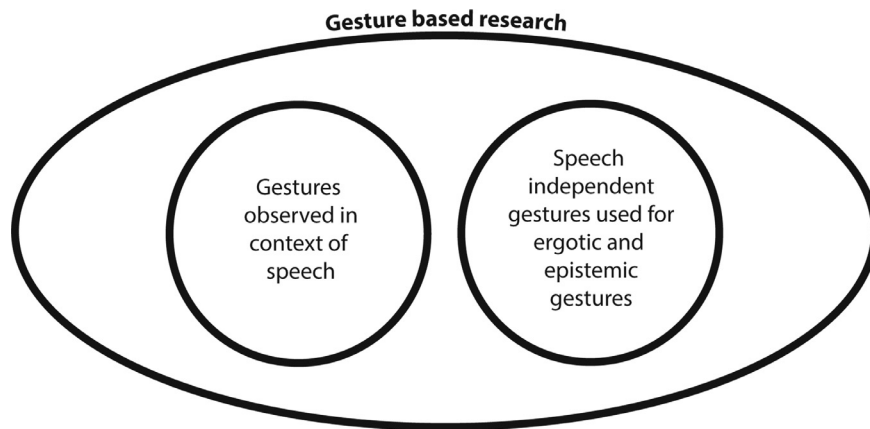


Fig. 9. Proposed expansion of gesture classification and typification.

interface. Section 5.2 has shown that some patterns can be identified by observing the type of technology used for a specific application e.g. VR was more likely to be used as a supportive technology for applications requiring 3D visualisation. However, there were no clear prescriptive patterns identified, or inherently advantageous technology-application combinations for a specific type of interfaces

Gesture selection is often influenced by the capabilities of the technologies facilitating their use (Schmidt, 2015). This may mean that gesture elicitation process may overlook the most appropriate gestures for a specific activity or interface. In other words, intuitiveness of gestures was not the primary decision factor for inclusion of gestures, which conflicts with the requirement for intuitive easy to use gestures ensuring effective implementation. Similarly, when two-dimensional (2D) surface gestures created by a number of gesture experts were compared to those created by end users, it was found that the users generated more diverse gestures than experts (Wobbrock et al., 2009). It was also found that even though users had no information on the source of gestures, they preferred gestures that were initially suggested by a larger number of users (Morris et al., 2010). Similar research has not yet been performed for 3D free-form gestures, but this finding may suggest the inclusion of large number of participants during gesture elicitation might be beneficial. As Huang et al. (2018) state discussing potential future work, studies are required that will identify the most natural and intuitive hand gestures, for particular interfaces. If these are, at least on a theoretical level, decoupled from the currently available technology and its capabilities, they can provide a useful knowledge to guide future interface and technology development.

9. Conclusion

Hand gestures were used extensively in interfaces for a variety of applications, facilitated by a large number of technologies. The applications fall under larger groupings of 3D modelling, assistive applications, data input, manipulation or navigation, and touchless interaction and control. Technologies can roughly be divided into visual based sensors and cameras, and physical wearables. The former are unobtrusive, and provide opportunity for natural movement, while the latter are generally more accurate and easier to set up and track, but can encumber the wearer. Depth cameras and accelerometers were the most frequently used technologies. However, sensors and cameras are gaining popularity as they provide the ability to implement interfaces requiring no physical contact with the user, which are more intuitive and less invasive. If accuracy and reliability were improved, these could become the primary technology for gesture interfaces in the future. Free-form gestures were often correlated with 3D application, and deictic gestures with interaction with different display types and robot control. HMM algorithm and gesture libraries and SDKs were used across majority of applications and facilitating technologies. DTM, FSM

and SVM algorithms were used only in applications supported by visual based technologies. However, overall there were no clear leaders or prescribed approaches in the technology, gestures or recognition methods used for different applications yet, however a variety of combinations of these show promise.

Since 2013, an expansion in number of articles exploring gesture-based interfaces is evident, in part due to technology development. However, it also seems that gestures, their nature, applicability, fitness for purpose and effectiveness were often not the primary focus in the articles reporting on the prototypes developed. Gestures were sole or one of the modes of interaction, but neither of the modes seemed to incorporate roles gestures typically serve in communication into the interfaces. For example, in multimodal applications, gestures were sometimes accompanied by speech. In these cases, they were rarely used to complement the speech, as would be expected based on their contextual roles. Instead, gestures and speech performed the same functions, and the choice to use gesture or speech for a certain task seemed to be primarily based on the ease of technical implementation. If this was not the case, speech was used for data input or a small number of predefined commands supplemented with a pointing gesture. In addition to this, the theory focusing on gesture research focuses on gestures in the context of verbal communication, while the applications being developed in majority of cases use different modalities for the interfaces. While this is largely due to different goals two approaches have, and motivations behind them, the fact that a gap between gesture theory and gesture application in interfaces exists remains. This may indicate that more research is needed to determine which gestures and gesture classes are the most appropriate for specific applications and technologies, potentially with a view of providing more structure for gesture elicitation process for interaction interfaces. A framework for gesture elicitation could guide future gesture based interaction interface development, ensuring important elements are considered in the process. Finally, observing gestures decoupled from speech may lead to new findings, which might contribute to gesture research, currently focusing on gestures when used simultaneously with speech.

Conflict of interests

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Tijana Vuletic, Alex Duffy, Laura Hay, Chris McTeague, Gerard Campbell, Madeleine Grealy

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijhcs.2019.03.011.

References

- Adamo-Villani, N., Heisler, J., Arns, L., 2007. Two gesture recognition systems for immersive math education of the deaf. In: Proceedings of the First International Conference on Immersive Telecommunications. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 9.
- Adeen, H.S., Atia, A., Amin, A., Victor, A., Essam, A., Gharib, E., Hussien, M., 2015. RemoAct: portable projected interface with hand gesture interaction. *J. Comput. Sci.* 11, 738–749.
- Adhikarla, V.K., Jakus, G., Sodnik, J., 2015. Design and evaluation of freehand gesture interaction for light field display. In: Seventeenth International Conference on Human-Computer Interaction, HCI International 2015, August 2, 2015 – August 7, 2015. Los Angeles, CA, United States. Springer Verlag, pp. 54–65.
- Al-Sayegh, R., Makatsoris, C., 2015. Vision-augmented molecular dynamics simulation of nanoindentation. *J. Nanomater.* 2015, 1–11.
- Al-Shamayleh, A.S., Ahmad, R., Abushariah, M.A., Alam, K.A., Jomhari, N., 2018. A systematic literature review on vision based gesture recognition techniques. *Multimed. Tools Appl.* 77, 1–64.
- Alvarez-Santos, V., Iglesias, R., Pardo, X.M., Regueiro, C.V., Canedo-Rodriguez, A., 2014. Gesture-based interaction with voice feedback for a tour-guide robot. *J. Visual Commun. Image Representation* 25, 499–509.
- Alves, R., Negrier, A., Sousa, L., Rodrigues, J.M.F., Felisberto, P., Gomes, M., Bica, P., 2015. Interactive 180° rear projection public relations. *Proc. Comput. Sci.* 51, 592–601.
- Ammar, C., Georgi, M., Schultz, T., 2014. Airwriting: a wearable handwriting recognition system. *Personal and Ubiquitous Computing* 18 (1), 191–203.
- Arroyave-Tobón, S., Osorio-Gómez, G., Cardona-McCormick, J.F., 2015. Air-modelling: a tool for gesture-based solid modelling in context during early design stages in AR environments. *Comput. Ind.* 66, 73–81.
- Asadzadeh, P., Kulik, L., Tanin, E., 2012. Gesture recognition using RFID technology. *Pers. Ubiquitous Comput.* 16, 225–234.
- Bai, H., Gao, L., El-Sana, J., Billingham, M., 2013. Free-hand interaction for handheld augmented reality using an RGB-depth camera. In: SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, SA 2013, November 19, 2013 – November 22, 2013. Hong Kong. Hong Kong Association for Computing Machinery, ACM Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH); ACM Special Interest Group on Computer-Human Interaction (SIGCHI).
- Bannach, D., Amft, O., Kunze, K.S., Heinz, E.A., Troster, G., Lukowicz, P., 2007. Waving real hand gestures recorded by wearable motion sensors to a virtual car and driver in a mixed-reality parking game. In: IEEE Symposium on Computational Intelligence and Games, 2007. CIG 2007. IEEE, pp. 32–39.
- Baudel, T., Beaudouin-Lafon, M., 1993. Charade: remote control of objects using free-hand gestures. *Commun. ACM* 36, 28–35.
- Beyer, G., Meier, M., 2011. Music interfaces for novice users: composing music on a public display with hand gestures. In: NIME, pp. 507–510.
- Bhuiyan, M., Picking, R., 2011. A gesture controlled user interface for inclusive design and evaluative study of its usability. *J. Softw. Eng. Appl.* 4, 513.
- Billingham, M., 2018. Chapter 14: gesture based interaction. In: Buxton, B. (Ed.), *Human Input to Computer Systems: Theories, Techniques and Technology*, Unpublished.
- Boboc, R.G., Dumitru, A.I., Antonya, C., 2015. Point-and-command paradigm for interaction with assistive robots. *Int. J. Adv. Rob. Syst.* 12, 75.
- Bolt, R.A., 1980. “Put-that-there”: Voice and Gesture at the Graphics Interface. ACM.
- Boulabiar, M.I., Burger, T., Poirier, F. & Coppin, G. 2011. A low-cost natural user interaction based on a camera hand-gestures recognizer. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Bourdot, P., Convard, T., Picon, F., Ammi, M., Touraine, D., Vézien, J.M., 2010. VR-CAD integration: multimodal immersive interaction and advanced haptic paradigms for implicit edition of CAD models. *Comput. Aided Des.* 42, 445–461.
- Braun, A., Zander-Walz, S., Majewski, M., Kuijper, A., 2017. Curved-free-form interaction using capacitive proximity sensors. *Proc. Comput. Sci.* 109, 59–66.
- Buchmann, V., Violich, S., Billingham, M., Cockburn, A., 2004. FingARTips: gesture based direct manipulation in Augmented Reality. In: Proceedings of the Second International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia. ACM, pp. 212–221.
- Buddhikot, A.G., Kulkarni, N.M., Shaligram, A.D., 2018. Hand gesture interface based on skin detection technique for automotive infotainment system. *Int. J. Image, Graph. Signal Process.* 10, 10.
- Cadoz, C., 1994. Les Réalités Virtuelles. Flammarion.
- Carbini, S., Delphin-Poulat, L., Perron, L., Viallet, J.E., 2006. From a wizard of Oz experiment to a real time speech and gesture multimodal interface. *Signal Process.* 86, 3559–3577.
- Carreira, M., Ting, K.L.H., Csobanka, P., Gonçalves, D., 2017. Evaluation of in-air hand gestures interaction for older people. *Univers. Access Inf. Soc.* 16, 561–580.
- Casasanto, D., Lozano, S., 2007. The meaning of metaphorical gestures. *Metaphor and Gesture*. Gesture Studies. John Benjamins Publishing, Amsterdam, the Netherlands (accessed 9.12.2012).
- Cash, P.J., 2018. Developing theory-driven design research. *Des. Stud.* 56, 84–119.
- Cassell, J., 1998. A framework for gesture generation and interpretation. In: Cipolla, R., Pentland, A. (Eds.), *Computer Vision in Human-Machine Interaction*. New York: Cambridge University Press, pp. 191–215.
- Cauchard, J.R., Fraser, M., Han, T., Subramanian, S., 2012. Steerable projection: exploring alignment in interactive mobile displays. *Pers. Ubiquitous Comput.* 16, 27–37.
- Cha, T., Maier, S., 2012. Eye gaze assisted human-computer interaction in a hand gesture controlled multi-display environment. In: Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction. ACM, pp. 13.
- Chen, M.-Y., Mummert, L., Pillai, P., Hauptmann, A., Suktharankar, R., 2010. Controlling your TV with gestures. In: Proceedings of the International Conference on Multimedia Information Retrieval. ACM, pp. 405–408.
- Chen, M., Alregib, G., Juang, B.H., 2011. An integrated framework for universal motion control. In: Proceedings of VRCAI 2011: ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications to Industry, pp. 513–518.
- Choi, E.H.C., Taib, R., Shi, Y., Chen, F., 2007. Multimodal user interface for traffic incident management in control room. *IET Intell. Transp. Syst.* 1, 27–36.
- Chu, C.-C.P., Dani, T.H., Gadh, R., 1997. Multi-sensory user interface for a virtual-reality-based computeraided design system. *Comput.-Aided Des.* 29, 709–725.
- Church, R.B., Goldin-Meadow, S., 1986. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23, 43–71.
- Cicirelli, G., Attolico, C., Guaragnella, C., D’Orazio, T., 2015. A kinetic-based gesture recognition approach for a natural human robot interface. *Int. J. Adv. Rob. Syst.* 12, 22.
- Colaço, A., Kirmani, A., Yang, H.S., Gong, N.-W., Schmandt, C., Goyal, V.K., 2013. Mime: compact, low power 3D gesture sensing for interaction with head mounted displays. In: Proceedings of the Twenty-Sixth Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 227–236.
- Coton, J., Veytizou, J., Thomann, G., Villeneuve, F., 2016. Feasibility Study of Hand Motion Analysis by the Leap Motion Sensor. Feasibility Study of Hand Motion Analysis by the Leap Motion Sensor.
- Covarrubias, M., Bordegoni, M., Cugini, U., 2015. A hand gestural interaction system for handling a desktop haptic strip for shape rendering. *Sens. Actuators A: Phys.* 233, 500–511.
- Dani, T.H., Gadh, R., 1997. Creation of concept shape designs via a virtual reality interface. *Comput.-Aided Des.* 29, 555–563.
- Dardas, N.H., Silva, J.M., El Saddik, A., 2014. Target-shooting exergame with a hand gesture control. *Multimed. Tools Appl.* 70, 2211–2233.
- Dave, D., Chowriappa, A., Kesavadas, T., 2013. Gesture interface for 3d cad modeling using kinect. *Computer-Aided Design and Applications* 10 (4), 663–669.
- Deller, M., Ebert, A., Bender, M., Hagen, H., 2006. Flexible gesture recognition for immersive virtual environments. In: Tenth International Conference on Information Visualization, 2006. IV 2006. IEEE, pp. 563–568.
- Denkowski, M., Dmitruk, K., Sadkowski, L., 2015. Building automation control system driven by gestures. *IFAC-Pap. OnLine* 48, 246–251.
- Dinh, D.-L., Kim, J.T., Kim, T.-S., 2014. Hand gesture recognition and interface via a depth imaging sensor for smart home appliances. *Energy Proc.* 62, 576–582.
- Djukic, T., Mandic, V., Filipovic, N., 2013. Virtual reality aided visualization of fluid flow simulations with application in medical education and diagnostics. *Comput. Biol. Med.* 43, 2046–2052.
- Dondi, P., Lombardi, L., Rocca, I., Malagodi, M., Licchelli, M., 2018. Multimodal workflow for the creation of interactive presentations of 360 spin images of historical violins. *Multimed. Tools Appl.* 77, 1–24.
- Foehrenbach, S., König, W.A., Gerken, J., Reiterer, H., 2009. Tactile feedback enhanced hand gesture interaction at large, high-resolution displays. *J. Visual Lang. Comput.* 20, 341–351.
- Fong, T.W., Conti, F., Grange, S., Baur, C., 2001. Novel interfaces for remote driving: gesture, haptic, and PDA. *Intelligent systems and smart manufacturing. Int. Soc. Opt. Photonics Vol.* 4195, 300–311.
- Freedman, N., 1972. The analysis of movement behavior during the clinical interview. *Stud. Dyadic Commun.* 7, 153–175.
- Fuhrmann, F., Kaiser, R., 2014. Multimodal interaction for future control centers: an interactive demonstrator. In: Sixteenth ACM International Conference on Multimodal Interaction, ICMI 2014, November 12, 2014 – November 16, 2014. Istanbul, Turkey. Association for Computing Machinery, Inc, pp. 66–67.
- Gangman, Y., Yen, N.Y., 2014. Development of highly interactive service platform for social learning via ubiquitous media. *Math. Prob. Eng.* 2014, 1–8.
- Garzotto, F., Valoriani, M., 2013. Touchless gestural interaction with small displays: a case study. In: Tenth ACM SIGCHI Italy Chapter Biannual Conference, CH Italy 2013, September 16, 2013 – September 20, 2013. Trento, Italy. Association for Computing Machinery, EIT ICT Labs; EU Project TERENCE; Noldus Information Technology; SRLabs.
- Gil, P., Mateo, C., Torres, F., 2014. 3D visual sensing of the human hand for the remote operation of a robotic hand. *Int. J. Adv. Rob. Syst.* 11, 26.
- Guerra-Casanova, J., Sánchez-Ávila, C., Bailador, G., De Santos Sierra, A., 2012. Authentication in mobile devices through hand gesture recognition. *Int. J. Inf. Secur.* 11, 65–83.
- Guiard, Y., 1987. Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. *J. Motor Behav.* 19, 486–517.
- Han, Y.C., Han, B.-J., 2014. Virtual pottery: a virtual 3D audiovisual interface using natural hand motions. *Multimed. Tools Appl.* 73, 917–933.
- Hasan, M.M., Mishra, P.K., 2012. Hand gesture modeling and recognition using geometric features: a review. *Can. J. Image Process. Comput. Vision* 3, 12–26.
- Hasanuzzaman, M., Zhang, T., Ampornaramveth, V., Ueno, H., 2006. Gesture-based

- human-robot interaction using a knowledge-based software platform. *Ind. Rob.* 33, 37–49.
- He, Z., Jin, L., Zhen, L., Huang, J., 2008. Gesture recognition based on 3D accelerometer for cell phones interaction. In: *IEEE Asia Pacific Conference on Circuits and Systems*, 2008. APCCAS 2008. IEEE, pp. 217–220.
- Hernoux, F., Christmann, O., 2015. A seamless solution for 3D real-time interaction: design and evaluation. *Virtual Reality* 19, 1–20.
- Holler, J., Beattie, G., 2003. How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process. *SEMIOTICA-LA HAYE THEN BERLIN* 146, 81–116.
- Holz, C., Wilson, A., 2011. Data miming: inferring spatial object descriptions from human gesture. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 811–820.
- Hoste, L., Signer, B., 2013. SpeeG2: a speech- and gesture-based interface for efficient controller-free text input. In: *2013 Fifteenth ACM International Conference on Multimodal Interaction, ICMi 2013, December 9, 2013 – December 13, 2013. Sydney, NSW, Australia. Association for Computing Machinery*, pp. 213–220.
- Huang, J., Jaiswal, P., Rai, R., 2018. Gesture-based system for next generation natural and intuitive interfaces. *AI EDAM* 33, 1–15.
- Huang, L., Zhou, Z., Liu, R., 2014. Research on Interaction-oriented gesture recognition. *Sens. Transducers* 163, 316–322.
- Hürst, W., Van Wezel, C., 2013. Gesture-based interaction via finger tracking for mobile augmented reality. *Multimed. Tools Appl.* 62, 233–258.
- Jacob, M.G., Wachs, J.P., 2014. Context-based hand gesture recognition for the operating room. *Pattern Recognit. Lett.* 36, 196–203.
- Jahani, H., Kavakli, M., 2018. Exploring a user-defined gesture vocabulary for descriptive mid-air interactions. *Cognit. Technol. Work* 20, 11–22.
- Kajastila, R., Lokki, T., 2013. Eyes-free interaction with free-hand gestures and auditory menus. *Int. J. Hum. Comput. Stud.* 71, 627–640.
- Kang, J., Zhong, K., Qin, S., Wang, H., Wright, D., 2013. Instant 3D design concept generation and visualization by real-time hand gesture recognition. *Comput. Ind.* 64, 785–797.
- Kapuscinski, T., Oszust, M., Wysocki, M., Warchol, D., 2015. Recognition of hand gestures observed by depth cameras. *International Journal of Advanced Robotic Systems* 12 (4), 36.
- Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Sergio Di, M., 2006. Accelerometer-based gesture control for a design environment. *Pers. Ubiquitous Comput.* 10, 285.
- Kendon, A., 1985. Some uses of gestures. In: Tannen, D., Saville-Troie, M. (Eds.), *Perspectives on Silence*. Norwood, NJ: Ablex, pp. 215–234.
- Kendon, A., 1986. Some reasons for studying gesture. *Semiotica* 62, 3–28.
- Kendon, A., 1988. How gestures can become like words. In: Poyatos, F. (Ed.), *Cross-cultural perspectives in nonverbal communication*. Toronto: Hogrefe, pp. 131–141.
- Kim, H., Albuquerque, G., Havemann, S., Fellner, D.W., 2005. Tangible 3D: hand gesture interaction for immersive 3D modeling. In: *IP/EGVE*, pp. 191–199.
- Kim, J., Mastnik, S., André, E., 2008. EMG-based hand gesture recognition for realtime biosignal interfacing. In: *Proceedings of the Thirteenth International Conference on Intelligent User Interfaces*. ACM, pp. 30–39.
- Kim, K., Kim, J., Choi, J., Kim, J., Lee, S., 2015. Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions. *Sensors (Switzerland)* 15, 1022–1046.
- Kim, M., Lee, J.Y., 2016. Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality. *Multimed. Tools Appl.* 75, 16529–16550.
- Kim, S.K., Kirchner, E.A., Stefes, A., Kirchner, F., 2017. Intrinsic interactive reinforcement learning—Using error-related potentials for real world human-robot interaction. *Sci. Rep.* 7, 17562.
- Kim, Y., Park, J., 2014. Study on interaction-induced symptoms with respect to virtual grasping and manipulation. *Int. J. Hum. Comput. Stud.* 72, 141–153.
- Kinect, M., 2018. Available: <https://msdn.microsoft.com/en-gb/library/hh438998.aspx> (Accessed 13.02.2018).
- Kratz, L., Smith, M., Lee, F., Wiizards, J., 2007. 3D gesture recognition for game play input. In: *Proceedings of the 2007 Conference on Future Play*. ACM, pp. 209–212.
- Krum, D.M., Omotoso, O., Ribarsky, W., Starner, T., Hodges, L.F., 2002. Speech and Gesture Multimodal Control of a Whole Earth 3D visualization Environment. Georgia Institute of Technology.
- Lauber, F., Bottcher, C., Butz, A., 2014. You've got the look: visualizing infotainment shortcuts in head-mounted displays. In: *Sixth International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2014, September 17, 2014 – September 19, 2014. Seattle, WA, United States. Association for Computing Machinery, Inc, American Honda Motor Co., Inc., Intel Corporation; Microsoft Corporation; Toyota Collaborative Safety Research Center (CSRC); University of Washington, College of Engineering et al.*
- LEAP MOTION INC., 2018. *LEAP MOTION* [Online]. Available: <https://www.leapmotion.com/> (Accessed 13.02.2018).
- Lee, E., Kiel, H., Dedenbach, S., Grill, I., Karrer, T., Wolf, M., Borchers, J.I., 2006. Symphony: an adaptive interactive orchestral conducting system for digital audio and video streams. In: *Conference on Human Factors in Computing Systems – Proceedings*, pp. 259–262.
- Lee, G., Shin, D., Shin, D., 2016. Mouse operation on monitor by interactive analysis of intuitive hand motions. *Multimed. Tools Appl.* 75, 15261–15274.
- Lin, J., Nishino, H., Kagawa, T., Utsumiya, K., 2012. A method of two-handed gesture interactions with applications based on commodity devices. *Comput. Math. Appl.* 63, 448–457.
- Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V., 2009. uWave: accelerometer-based personalized gesture recognition and its applications. *Pervasive Mob. Comput.* 5, 657–675.
- Liu, W., Fan, Y., Li, Z., Zhang, Z., 2015. RGBD video based human hand trajectory tracking and gesture recognition system. *Mathematical Problems in Engineering*, 2015.
- Liu, Y., Wang, X., Yan, K., 2018. Hand gesture recognition based on concentric circular scan lines and weighted K-nearest neighbor algorithm. *Multimed. Tools Appl.* 77, 209–223.
- Lopes, D.S., De Figueiredo Parreira, P.D., Paulo, S.F., Nunes, V., Rego, P.A., Neves, M.C., Rodrigues, P.S., Jorge, J.A., 2017. On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface. *J. Biomed. Inf.* 72, 140–149.
- Lu, G., Shark, L.-K., Hall, G., Zeshan, U., 2012. Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality* 16, 243–252.
- Lv, Z., Halawani, A., Feng, S., Ur Rehman, S., Li, H., 2015. Touch-less interactive augmented reality game on vision-based wearable device. *Pers. Ubiquitous Comput.* 19, 551–567.
- Ma, C., Zhang, Y., Wang, A., Wang, Y., Chen, G., 2018. Traffic command gesture recognition for virtual urban scenes based on a spatiotemporal convolution neural network. *ISPRS Int. J. Geo-Inf.* 7, 37.
- Mahr, A., Endres, C., Muller, C., Schneeberger, T., 2011. Determining human-centered parameters of ergonomic micro-gesture interaction for drivers using the theater approach. In: *Third International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2011, November 30, 2011 – December 2, 2011. Salzburg, Austria. Association for Computing Machinery*, pp. 151–157.
- Marasovic, T., Papic, V., Marasovic, J., 2015. Motion-based gesture recognition algorithms for robot manipulation. *Int. J. Adv. Rob. Syst.* 12, 51.
- McNeill, D., 1985. So you think gestures are nonverbal? *Psychol. Rev.* 92 (3), 350.
- McNeill, D., 1987. *Psycholinguistics: A new Approach*. Harper & Row Publishers.
- McNeill, D., 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago press.
- McNeill, D., 2006. Gesture: a psycholinguistic approach. *The Encyclopedia of Language and Linguistics*. pp. 58–66.
- Memo, A., Zanuttigh, P., 2018. Head-mounted gesture controlled interface for human-computer interaction. *Multimed. Tools Appl.* 77, 27–53.
- Milani, F., De Marchi, A.C.B., Rieder, R., 2017. Usability guidelines to develop gesture-based serious games for health: a systematic review. In: *2017 Nineteenth Symposium on Virtual and Augmented Reality (SVR)*. IEEE, pp. 188–194.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6, e1000097.
- Morris, M.R., Wobbrock, J.O., Wilson, A.D., 2010. Understanding users' preferences for surface gestures. In: *Proceedings of Graphics Interface 2010. Canadian Information Processing Society*, pp. 261–268.
- Moustakas, K., Nikolakis, G., Tzovaras, D., Carhini, S., Bernier, O., Viallet, J.E., 2009. 3D content-based search using sketches. *Pers. Ubiquitous Comput.* 13, 59–67.
- Nazemi, K., Burkhardt, D., Stab, C., Breyer, M., Wichert, R., Fellner, D.W., 2011. Natural Gesture Interaction with Accelerometer-Based Devices in Ambient Assisted Environments.
- New, J.R., Hasanbelliu, E., Aguilar, M., 2003. Facilitating user interaction with complex systems via hand gesture recognition. In: *Proceedings of the 2003 Southeastern ACM Conference*. Savannah, GA.
- Ni, T., Bowman, D.A., North, C., McMahan, R.P., 2011. Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *Int. J. Hum. Comput. Stud.* 69, 551–562.
- Nicola, S., Handrea, F.-L., Crişan-Vida, M., Stoicu-Tivadar, L., 2017. DNA encoding training using 3D gesture interaction. *The Practice of Patient Centered Care: Empowering and Engaging Patients in the Digital Era* 244, pp. 63.
- Niezen, G., Hancke, G.P., 2008. Gesture recognition as ubiquitous input for mobile phones. In: *International Workshop on Devices that Alter Perception (DAP 2008)*, in Conjunction with Ubicomp. Citeseer, pp. 17–21.
- O'hagan, R., Zelinsky, A., Rougeaux, S., 2002. Visual gesture interfaces for virtual environments. *Interact. Comput.* 14, 231–250.
- O'connor, T.F., Fach, M.E., Miller, R., Root, S.E., Mercier, P.P., Lipomi, D.J., 2017. The language of glove: wireless gesture decoder with low-power and stretchable hybrid electronics. *PLoS One* 12, e0179766.
- Ojeda-Castelo, J.J., Piedra-Fernandez, J.A., Iribarne, L., Bernal-Bravo, C., 2018. KINEET: application for learning and rehabilitation in special educational needs. *Multimed. Tools Appl.* 77, 1–27.
- Osti, F., Ceruti, A., Liverani, A., Caligiana, G., 2017. Semi-automatic design for disassembly strategy planning: an augmented reality approach. *Proc. Manuf.* 11, 1481–1488.
- Palacios, J.M., Sagües, C., Montijano, E., Llorente, S., 2013. Human-computer interaction based on hand gestures using RGB-D sensors. *Sensors* 13, 11842–11860.
- Pan, G., Wu, J., Zhang, D., Wu, Z., Yang, Y., Li, S., 2010. GeeAir: a universal multimodal remote control device for home appliances. *Pers. Ubiquitous Comput.* 14, 723–735.
- Pang, Y.Y., Ismail, N.A., Siang Gilbert, P.L., 2010. A real time vision-based hand gesture interaction. In: *Asia Modelling Symposium 2010: Fourth International Conference on Mathematical Modelling and Computer Simulation, AMS2010, May 26, 2010 – May 28, 2010. Kota Kinabalu, Borneo, Malaysia. IEEE Computer Society*, pp. 237–242.
- Park, K.-B., Lee, J.Y., 2018. New design and comparative analysis of smartwatch metaphor-based hand gestures for 3D navigation in mobile virtual reality. *Multimed. Tools Appl.* 1–21. <https://link.springer.com/article/10.1007/s11042-018-6403-9>.
- Pisharady, P.K., Saerbeck, M., 2015. Recent methods and databases in vision-based hand gesture recognition: a review. *Comput. Vision Image Understanding* 141, 152–165.
- Pook, P.K., Ballard, D.H., 1996. Deictic human/robot interaction. *Rob. Autom. Syst.* 18, 259–269.

- Poupyrev, I., Maruyama, S., Rekimoto, J., 2002. Ambient touch: designing tactile interfaces for handheld devices. In: *Proceedings of the Fifteenth Annual ACM Symposium on User Interface Software and Technology*. ACM, pp. 51–60.
- Qin, S., Wright, D.K., Kang, J., Prieto, P.A., 2006. Use of three-dimensional body motion to free-form surface design. *Proc. Inst. Mech. Eng.* 220, 335–339.
- Quek, F., 2004. The catchment feature model: a device for multimodal fusion and a bridge between signal and sense. *EURASIP J. Appl. Signal Process.* 2004, 1619–1636.
- Quek, F.K.H., 1995. Eyes in the interface. *Image Vision Comput.* 13, 511–525.
- Rautaray, S.S., Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artif. Intell. Rev.* 43, 1–54.
- Reale, M.J., Canavan, S., Yin, L., Hu, K., Hung, T., 2011. A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing. In: *Special Issue Dedicated to the 2010 IEEE International Conference on Multimedia and EXPO*. Institute of Electrical and Electronics Engineers Inc., pp. 474–486.
- Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., Rigoll, G., 2007. Static and dynamic hand-gesture recognition for augmented reality applications. *Human-Computer Interaction*. 728–737.
- Rekimoto, J., 2001. Gesture wrist and gesture pad: unobtrusive wearable interaction devices. In: *Fifth International Symposium on Wearable Computers*, 2001. *Proceedings*. IEEE, pp. 21–27.
- Riduwan, M., Basori, A.H., Mohamed, F., 2013. Finger-based gestural interaction for exploration of 3D heart visualization. *Proc. Soc. Behav. Sci.* 97, 684–690.
- Riener, A., Ferscha, A., Bachmair, F., Hagmuller, P., Lemme, A., Muttenthaler, D., Puhlinger, D., Rogner, H., Tappe, A., Weger, F., 2013. Standardization of the in-car gesture interaction space. In: *Fifth International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, *AutomotiveUI 2013*, October 27, 2013 – October 30, 2013. Eindhoven, Netherlands. Association for Computing Machinery, pp. 14–21.
- Rimé, B., Schiaratura, L., 1991. Gesture and speech. In: Feldman, R.S., Rimé, B. (Eds.), *Studies in emotion & social interaction. Fundamentals of nonverbal behavior*. New York, NY, US: Cambridge University Press; Paris, pp. 239–281 France: Editions de la Maison des Sciences de l'Homme.
- Robinson, G., Ritchie, J.M., Day, P.N., Dewar, R.G., 2007. System design and user evaluation of Co-Star: an immersive stereoscopic system for cable harness design. *Comput. Aided Des.* 39, 245–257.
- Roccetti, M., Marfia, G., Semeraro, A., 2012. Playing into the wild: a gesture-based interface for gaming in public spaces. *J. Visual Commun. Image Represent.* 23, 426–440.
- Rodrigues, É., Carreira, M., Gonçalves, D., 2014. Developing a multimodal interface for the elderly. *Proc. Comput. Sci.* 27, 359–368.
- Rogalla, O., Ehrenmann, M., Zollner, R., Becher, R., Dillmann, R., 2002. Using gesture and speech control for commanding a robot assistant. In: *Eleventh IEEE International Workshop on Robot and Human Interactive Communication*, 2002. *Proceedings*. IEEE, pp. 454–459.
- Rossol, N., Cheng, I., Rui, S., Basu, A., 2014. Touchfree medical interfaces. In: *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2014, pp. 6597–6600.
- Ruppert, G.C.S., Reis, L.O., Amorim, P.H.J., De Moraes, T.F., Da Silva, J.V.L., 2012. Touchless gesture user interface for interactive image visualization in urological surgery. *World J. Urol.* 30, 687–691.
- Santos, C.A., Saleme, E.B., De Andrade, J.C., 2015. A systematic review of data exchange formats in advanced interaction environments. *Development* 10, 123–140.
- Santos, B.S., Cardoso, J., Ferreira, B.Q., Ferreira, C., Dias, P., 2016. Developing 3D freehand gesture-based interaction methods for virtual walkthroughs: using an iterative approach. *Handbook of Research on Human-Computer Interfaces, Developments, and Applications*. IGI Global.
- Santos, D.G., Fernandes, B.J.T., Bezerra, B.L.D., 2015. HAGR-D: a novel approach for gesture recognition with depth maps. *Sensors* 15, 28646–28664.
- Savage-Carmona, J., Billingham, M., Holden, A., 1998. The VirBot: a virtual reality robot driven with multimodal commands. *Expert Syst. Appl.* 15, 413–419.
- Saxen, F., Rashid, O., Al-Hamadi, A., Adler, S., Kernchen, A., Mecke, R., 2014. Image-based methods for interaction with head-worn worker-assistance systems. *J. Intell. Learn. Syst. Appl.* 6, 141–152.
- Schmidt, A., 2015. Following or leading?: the HCI community and new interaction technologies. *Interactions* 22, 74–77.
- Schreiber, M., Von Wilamowitz-Moellendorf, M., Bruder, R., 2009. New interaction concepts by using the wii remote. In: *Thirteenth International Conference on Human-Computer Interaction, HCI International 2009*, July 19, 2009 – July 24, 2009. San Diego, CA, United States. Springer Verlag, pp. 261–270.
- Shapir, O., Goldschmidt, G., Yezioro, A., 2007. Conceptual design: an operational prescription for a computer support system. In: *Computer Graphics, Imaging and Visualisation*, 2007. CGIV'07. IEEE, pp. 513–521.
- Shim, J., Yang, Y., Kang, N., Seo, J., Han, T.-D., 2016. Gesture-based interactive augmented reality content authoring system using HMD. *Virtual Reality* 20, 57–69.
- Sodhi, R., Poupyrev, I., Glisson, M., Israr, A., 2013. AIREAL: interactive tactile experiences in free air. *ACM Trans. Graphics* 32, 134.
- Suarez, J., Murphy, R.R., 2012. Hand gesture recognition with depth images: a review. In: *Ro-man*, 2012 IEEE. IEEE, pp. 411–417.
- Takahashi, M., Fujii, M., Naemura, M., Satoh, S.I., 2013. Human gesture recognition system for TV viewing using time-of-flight camera. *Multimed. Tools Appl.* 62, 761–783.
- Togootogtokh, E., Shih, T.K., Kumara, W., Wu, S.-J., Sun, S.-W., Chang, H.-H., 2018. 3D finger tracking and recognition image processing for real-time music playing with depth sensors. *Multimed. Tools Appl.* 77, 9233–9248.
- Trigueiros, P., Ribeiro, F., Reis, L.P., 2015. Generic System for human-computer gesture interaction: applications on sign language recognition and robotic soccer refereeing. *J. Intell. Rob. Syst.* 80, 573–594.
- Van Den Bergh, M., Carton, D., De Nijs, R., Mitsou, N., Landsiedel, C., Kuehnlz, K., Wollherr, D., Van Gool, L., Buss, M., 2011. Real-time 3D hand gesture interaction with a robot for understanding directions from humans. In: *Twentieth IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2011*, July 31, 2011 – August 3, 2011. Atlanta, GA, United States. Institute of Electrical and Electronics Engineers Inc., pp. 357–362.
- Vinayak, Murugappan, S., Liu, H., Ramani, K., 2013. Shape-It-Up: hand gesture based creative expression of 3D shapes using intelligent generalized cylinders. *Comput. Aided Des.* 45, 277–287.
- Vinayak, Ramani, K., 2015. A gesture-free geometric approach for mid-air expression of design intent in 3D virtual pottery. *Comput. Aided Des.* 69, 11–24.
- Vosinakis, S., Koutsabasis, P., 2018. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality* 22, 47–62.
- Wachs, J.P., Kölsch, M., Stern, H., Edan, Y., 2011. Vision-based hand-gesture applications. *Commun. ACM* 54, 60–71.
- Wachs, J.P., Stern, H.I., Edan, Y., Gillam, M., Handler, J., Feied, C., Smith, M., 2008. A gesture-based tool for sterile browsing of radiology images. *J. Am. Med. Inf. Assoc.* 15, 321–323.
- Wagner, P., Malisz, Z., Kopp, S., 2014. Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232.
- Waldherr, S., Romero, R., Thrun, S., 2000. A gesture based interface for human-robot interaction. *Auton. Rob.* 9, 151–173.
- Wang, K., Xiao, B., Xia, J., Li, D., Luo, W., 2016. A real-time vision-based hand gesture interaction system for virtual EAST. *Fusion Eng. Des.* 112, 829–834.
- Widmer, A., Schaer, R., Markonis, D., Müller, H., 2014. Gesture interaction for content-based medical image retrieval. In: *2014 Fourth ACM International Conference on Multimedia Retrieval, ICMR 2014*, April 1, 2014 – April 4, 2014. Glasgow, United Kingdom. Association for Computing Machinery, pp. 503–506.
- Wilson, A., Oliver, N.G., 2003. Windows: robust stereo vision for gesture-based control of windows. In: *Proceedings of the Fifth International Conference on Multimodal Interfaces*. ACM, pp. 211–218.
- Wobbrock, J.O., Morris, M.R., Wilson, A.D., 2009. User-defined gestures for surface computing. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 1083–1092.
- Wright, M., Lin, C.-J., O'Neill, E., Cosker, D., Johnson, P., 2011. 3D gesture recognition: an evaluation of user and system performance. In: *International Conference on Pervasive Computing*. Springer, pp. 294–313.
- Wu, H., Wang, J., Zhang, X., 2016. User-centered gesture development in TV viewing environment. *Multimed. Tools Appl.* 75, 733–760.
- Xian, W., Tarrío Alonso, P., Bernardos Barbolá, A.M., Metola Moreno, E., Casar Correda, J.R., 2012. User-independent accelerometer-based gesture recognition for mobile devices. *Adv. Distrib. Comput. Artif. Intell. J.* 1, 11–25.
- Xiao, Y., Peng, Q., 2017. A hand gesture-based interface for design review using leap motion controller. In: *DS 87-8 Proceedings of the Twenty-First International Conference on Engineering Design (ICED 17) Vol 8: Human Behaviour in Design*, Vancouver, Canada, 21-25.08. 2017, pp. 239–248.
- Xie, Y., Xu, R., 2013. Natural bare-hand interaction for remote operating large touch screen. *J. Comput.* 8, 2512–2519.
- Xu, D., Wu, X., Chen, Y.-L., Xu, Y., 2015. Online dynamic gesture recognition for human robot interaction. *J. Intell. Rob. Syst.* 77, 583–596.
- Xu, Z., Xiang, C., Wang, W.H., Yang, J.H., Lantz, V., Wang, K.Q., 2009. Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. In: *International Conference on Intelligent User Interfaces*, *Proceedings IUI*, pp. 401–405.
- Yeo, H.-S., Lee, B.-G., Lim, H., 2015. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. *Multimed. Tools Appl.* 74, 2687–2715.
- Yoshioka, K., 2005. Linguistic and Gestural Introduction and Tracking of Referents in L1 and L2 Discourse.
- Yamada, M., Shan, J., Sakai, K., Murase, Y., Okabayashi, K., 2014. Immediately-available input method using one-handed motion in arbitrary postures. *Procedia Computer Science* 39, 51–58.
- Yuan, C., 2005. Visual tracking for seamless 3d interactions in augmented reality. In: *International Symposium on Visual Computing*. Springer, Berlin, Heidelberg, pp. 321–328.
- Zaii, I.-A., Pentiu, S.-G., Vatavu, R.-D., 2015. On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. *Pers. Ubiquitous Comput.* 19, 821–838.
- Zeng, W., Wang, C., Wang, Q., 2018. Hand gesture recognition using leap motion via deterministic learning. *Multimed. Tools Appl.* 77, 1–22.
- Zhou, Y., Cheng, Z., Jing, L., Wang, J., Huang, T., 2014. Pre-classification based hidden Markov model for quick and accurate gesture recognition using a finger-worn device. *Appl. Intell.* 40, 613–622.
- Zhu, C., Sheng, W., 2011. Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *IEEE Trans. Syst. Man Cybernet. Syst. Hum.* 41, 569–573.
- Zimmerman, T.G., Lanier, J., Blanchard, C., Bryson, S., Harvill, Y., 1987. A hand gesture interface device. In: *ACM SIGCHI Bulletin*. ACM, pp. 189–192.