School II – School of Computing Science, Business Administration, Economics and Law
Department of Computing Science

# Designing and Evaluating Scalable Mid-Air Gestures for Interactive Systems

The School of Computing Science, Business Administration, Economics and Law of the University of Oldenburg has accepted as partial fulfilment of the requirements for the degree and title of

Doctor of Science (Dr. rer. nat.)

by **M.Sc. Masoumehsadat Hosseini**
born on 22 December 1992 in Arak, Iran

Supervisor:

**Prof. Dr. Susanne Boll**

Other assessor:

**Prof. Dr. Kristof Van Laerhoven**

Date of thesis defence: 3 July 2026

# Eigenständigkeit

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und der Inhalt der Arbeit noch nicht für eine Diplom- oder ähnliche Prüfungsarbeit verwendet wurde.

Oldenburg, den 17. November 2025

Masoumehsadat Hosseini

# Abstract

As pervasive computing environments, such as smart homes, in-vehicle systems, and AR/VR applications, become increasingly widespread, the limitations of traditional input modalities, including keyboards, mice, touchscreens, and voice commands, are becoming more apparent. For instance, keyboards and mice are tied to static workstations and unsuitable for immersive or mobile settings; touchscreens depend on physical surfaces and thus cannot be used when hands are occupied, dirty, or gloved; and voice input, although convenient, is unreliable in noisy environments and often undesirable in shared or private contexts.

Mid-air gestural interaction represents a complementary and alternative input modality in situations where touch- or voice-based interaction is limited or unavailable. Gestures constitute a natural form of communication and can support interaction across multiple devices and settings. However, despite advances in sensing technologies and an expanding body of research, existing systems remain faced with challenges that limit their broader adoption. These include the lack of a consistent gesture vocabulary across application domains, insufficient understanding of how gestures are physically executed and vary across users, and limited ecological validity in studies that often neglect real-world contexts of use. Moreover, existing motion datasets primarily capture functional or biomedical movements, leaving communicative mid-air gestures in HCI underrepresented.

This dissertation investigates how to design and evaluate mid-air gestures as a reliable input modality across pervasive computing environments. Three central perspectives are addressed: (1) analyzing the consistency of gesture vocabularies across domains to assess the potential for a consensus-based gesture set, (2) exploring user preferences and motivations for integrating mid-air gesture use in different smart home contexts, and (3) to produce a publicly accessible dataset of mid-air gesture performances using multimodal motion capture, and to examine the physical execution of gestures to identify patterns of variability that inform robust gesture design and recognition.

To address these objectives, we employed a literature review, scenario-based elicitation studies, and developed a high-fidelity multimodal dataset for analysis of mid-air gesture performance. The findings reveal patterns of consistency across domains, highlight how user preferences are shaped by situational factors, and uncover variability in gesture performance across individuals and repetitions. This work contributes empirical evidence and methodological resources to inform the development of robust and generalizable gesture vocabularies, offering practical guidelines for creating intuitive, consistent, and transferable mid-air interaction techniques.

# Zusammenfassung

Mit der zunehmenden Verbreitung von allgegenwärtigen Computerumgebungen wie Smart Homes, Fahrzeugsystemen und AR/VR-Anwendungen werden die Grenzen traditioneller Eingabemethoden wie Tastaturen, Mäuse, Touchscreens und Sprachbefehle immer deutlicher. So sind Tastaturen und Mäuse an statische Arbeitsplätze gebunden und für immersive oder mobile Umgebungen ungeeignet; Touchscreens sind auf physische Oberflächen angewiesen und können daher nicht verwendet werden, wenn die Hände beschäftigt, schmutzig oder mit Handschuhen bekleidet sind; und die Spracheingabe ist zwar bequem, aber in lauten Umgebungen unzuverlässig und in gemeinsamen oder privaten Kontexten oft unerwünscht.

Die gestische Interaktion in der Luft ist eine alternative und ergänzende Eingabemodalität, wenn die berührungs- oder sprachbasierte Interaktion eingeschränkt oder nicht verfügbar ist. Gesten sind eine natürliche Form der Kommunikation, die die Interaktion über mehrere Geräte und Umgebungen hinweg unterstützen kann. Trotz der Fortschritte in der Sensortechnologie und einer wachsenden Zahl von Forschungsarbeiten stehen bestehende Systeme jedoch vor Herausforderungen, die ihre breitere Einführung einschränken. Zu diesen Herausforderungen gehören ein Mangel an einheitlichen Gestenvokabularen für verschiedene Anwendungsbereiche, ein unzureichendes Verständnis dafür, wie Gesten physisch ausgeführt werden und wie sie sich von Benutzer zu Benutzer unterscheiden, sowie eine begrenzte ökologische Validität in Studien, die oft die realen Anwendungskontexte vernachlässigen. Darüber hinaus erfassen bestehende Bewegungsdatensätze in erster Linie funktionelle oder biomedizinische Bewegungen, sodass kommunikative Gesten in der Mensch-Computer-Interaktion (HCI) unterrepräsentiert sind.

Diese Dissertation untersucht, wie Gesten in der Luft als zuverlässige Eingabemodalität in allgegenwärtigen Computerumgebungen gestaltet und bewertet werden können. Drei zentrale Perspektiven werden dabei behandelt: (1) Analyse der Konsistenz von Gestenvokabularen über verschiedene Domänen hinweg, um das Potenzial für einen konsensbasierten Gestensatz zu bewerten, (2) Untersuchung der Präferenzen und Motivationen der Nutzer für die Integration von Gestensteuerung in der Luft in verschiedenen Smart-Home-Kontexten und (3) Erstellung eines öffentlich zugänglichen Datensatzes von Gesten in der Luft unter Verwendung multimodaler Bewegungserfassung und Untersuchung der physischen Ausführung von Gesten, um Variabilitätsmuster zu identifizieren, die Aufschluss über ein robustes Gestendesign und eine robuste Gestenerkennung geben.

Um diese Ziele zu erreichen, haben wir eine Literaturrecherche und szenariobasierte Erhebungsstudien durchgeführt und einen hochauflösenden multimodalen Datensatz für die Analyse der Gestenleistung in der Luft entwickelt. Die Ergebnisse zeigen Muster der Konsistenz über verschiedene Bereiche hinweg, verdeutlichen, wie die Präferenzen der Nutzer durch situative Faktoren geprägt werden, und decken Variabilitäten in der Gestenausführung zwischen verschiedenen Personen

und Wiederholungen auf. Diese Arbeit liefert empirische Belege und methodische Ressourcen für die Entwicklung robuster und verallgemeinerbarer Gestenvokabulare und bietet praktische Leitlinien für die Schaffung intuitiver, konsistenter und übertragbarer Interaktionstechniken in der Luft.

# Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Susanne Boll, for her unwavering support and guidance throughout my PhD journey. Her dedication to research, her willingness to make time for me, even during her busiest periods, and enthusiasm have been a great source of inspiration and played a significant role in shaping this work. I am especially grateful for the outstanding research environment she created and for the freedom she gave me to explore novel ideas.

I would also like to thank Prof. Dr. Kristof Van Laerhoven, Prof. Dr. Andreas Peter, and Dr. Nadine Wagener for serving on my doctoral examination committee and for generously dedicating their time and expertise to the evaluation of this thesis.

Many thanks go to my colleagues at the University of Oldenburg and the OFFIS Institute for Information Technology, past and present, for fostering a supportive, collaborative, and inspiring research environment. I am deeply grateful to all of my co-authors and project partners for their valuable contributions and enjoyable collaborations. I also sincerely thank everyone with whom I had stimulating discussions at conferences, seminars, and workshops.

I'm also deeply grateful to my parents, my brother, and my sisters for their constant encouragement and support, which have meant the world to me.

# Contents

# 1 Introduction

In the following chapter, we present the idea behind this dissertation through framing the challenge of designing and understanding gesture-based interaction in devices, applications, and systems. Next, we introduce the presented work and the research questions that guided our investigations, and provide a breakdown of the contributions made. We further elaborate on our research approach and methodology, and conclude the chapter with a general overview.

## 1.1 Motivation and Problem Statement

The way humans interact with computers and computerized environments has undergone several fundamental changes in the last decades. In the early beginnings of computing, command line-based interfaces were used to issue instructions that had to be recalled by users, requiring deep knowledge of the available command set or tedious studying of manuals. With the introduction of graphical user interfaces in the 1980s, users were relieved of much of this burden. This was a transformative milestone that shifted the interaction paradigm from recall to visual recognition by leveraging graphical elements such as windows, icons, menus, and pointers (WIMP). This development enabled the manipulation of computers using pointing devices, such as the mouse, which significantly reduced cognitive load and made computing more accessible to non-expert users.

Beginning in the early 2000s, the GUI-centric paradigm gradually gave way to a touch-based interaction paradigm, which enables users to interact directly with devices using their fingers or a stylus, as the main new interaction principle. With the rise of these touch-sensitive surfaces, multi-finger gestures (e.g., pinch and zoom gestures) became prominent. We now use these gestures to zoom in and out on maps or photos on touch-enabled surfaces, such as mobile phones or tablet computers [Han05]. This new paradigm also required novel ways of organizing interfaces, relying much less on hierarchical menus, scroll bars, and dedicated representations of pointers, such as mouse pointers. Furthermore, the touch interaction paradigm is much more targeted toward mobile, contextualized devices and their respective interaction constraints. It is striking to see that today all major desktop operating systems have been designed to follow design insights gained from mobile interfaces, including multi-touch interaction and multi-application handling.

In hindsight, the transition between these two paradigms was difficult in multiple respects. Although the touch technology was established and available, its adoption in interactive systems still faced many challenges. First, a new touch vocabulary had to be established. While the properties of single touch were known from the first generation of touch-enabled mobile devices, such as the Apple Newton1 [Smi94], it was still unclear what multi-touch gestures could be used, which would allow for consistency, both internally within an application and externally across applications. New design guidelines had to be researched and developed to facilitate the broad adoption of the new paradigm across various applications and domains.

This dissertation is motivated by our conviction that we are facing similar problems at this very moment, as a new interaction paradigm shift appears to be necessary to overcome the limitations of 2D-interactive surfaces and the associated touch interaction paradigm. As computing systems become increasingly prevalent in everyday environments, such as smart homes or AR/VR systems, 2D touch surfaces are losing their importance. The emergence of wearable devices and sensor-based technologies further accelerates this shift, gradually replacing 2D gestures with more natural modalities, including voice and mid-air spatial gestures that leverage the whole body for interaction. While touch can be central to personal devices such as smartphones and tablets, it is inherently limited by its dependence on physical surfaces and can be unsuitable when hands are occupied, dirty, or gloved. Similarly, voice interaction, although increasingly popular, can be impractical in noisy environments or undesirable in shared or private contexts.

In this dissertation, we focus on mid-air gesture interaction, a specific form of kinaesthetic interaction, in which human body movements and postures are employed as input mechanisms for interacting with computer systems [FFK08]. Kinaesthetic interaction encompasses a wide range of body inputs, including hand gestures, facial expressions, gaze, foot pressure, and locomotion. Within this spectrum, mid-air interaction is commonly used to denote touchless, gesture-based interactions with remote displays or devices and especially in this dissertation, we define mid-air gesture interaction as comprising the following key attributes:

a. Touchless interaction with distant displays or devices.

b. Real-time tracking of the user's body (or specific parts such as hands) is typically achieved using wearable sensors or vision-based systems.

c. Body movements and gestures are identified and interpreted to infer user intent, goals, and interactions with digital content or devices.

Arguably, mid-air interaction, when robust and fit-for-purpose, may demonstrate great potential, either complementing other HCI styles or serving as a primary interaction style in many user scenarios, such as interactive VR/AR applications, human-robot interactions, and interactions with autonomous devices. Due to their ability to be performed eyes-free [HZKB17] and their perceived naturalness, intuitiveness, and ease of learning [BBL93; HLB$^+$10], mid-air (or free-hand) gestures have become a compelling and increasingly popular interaction paradigm. Additionally, the practical feasibility of such interaction is further reinforced by advances in sensing technologies, including depth cameras (e.g., Intel RealSense), radar systems (e.g., Google Soli), and electromagnetic trackers (e.g., Polhemus).

These devices now enable precise, real-time tracking of body movements, making gesture input viable for a wide range of applications.

Despite considerable progress in sensing technologies and growing interest in mid-air gesture interaction, several key challenges still prevent its widespread and effective deployment in real-world interactive systems. First, although a large number of gesture-based interaction studies have been conducted, most have been designed and targeted around isolated use cases, such as TV control, gaming, or surgical environments, and do not (necessarily) span across different devices, scenarios, or domains. As a result, similar actions may be mapped to different gestures depending on the context, or conversely, identical gestures may be used to trigger different functions across systems. This lack of consistency imposes a cognitive burden on users, who must effectively learn a new "gesture set" for each system they encounter [NKQK13]. Rather than supporting natural and seamless interaction, this fragmentation diminishes usability, increases error rates, and ultimately discourages users from adopting mid-air gestures as a reliable modality [VK19; DRSH18b].

Second, the performance and robustness of gestures remain poorly understood. There is natural variability in how users physically perform gestures and movements, both across individuals (interpersonal) and within the same individual across repetitions (intrapersonal). Without a systematic evaluation of gesture motion characteristics, it is difficult to determine which gestures are reliable and suitable for deployment in different real-world systems. Existing studies in HCI often capture users' intended gestures using images or descriptive information of gestures, rather than measuring their actual execution. As a result, these studies offer limited insight into the physical feasibility and ergonomic characteristics of gestures. Addressing these limitations requires a datasets that capture detailed motion data gestures in HCI across multiple users.

Third, ecological validity is rarely addressed in prior work. Many studies rely on abstract laboratory settings and narrowly defined single-device tasks, which limit insights into how users prefer to design and apply gestures in different scenarios in everyday contexts such as smart homes. Consequently, critical influences on gesture preferences, including scenario familiarity, environmental factors, and the affordances of devices, remain largely understudied.

Fourth, gesture research currently lacks comprehensive datasets that capture the richness and variability of communicative mid-air gestures in human–computer interaction. While some large-scale multimodal datasets exist, they are predominantly focused on biomechanical or functional hand movements, such as grasping, rehabilitation tasks, or sign language. Although valuable in biomedical and assistive contexts, these datasets do not necessarily reflect the communicative and interactive mid-air gestures commonly studied in HCI. This gap limits the ability of researchers and system designers to analyze execution patterns, assess gesture reliability, and benchmark recognition technologies across diverse users.

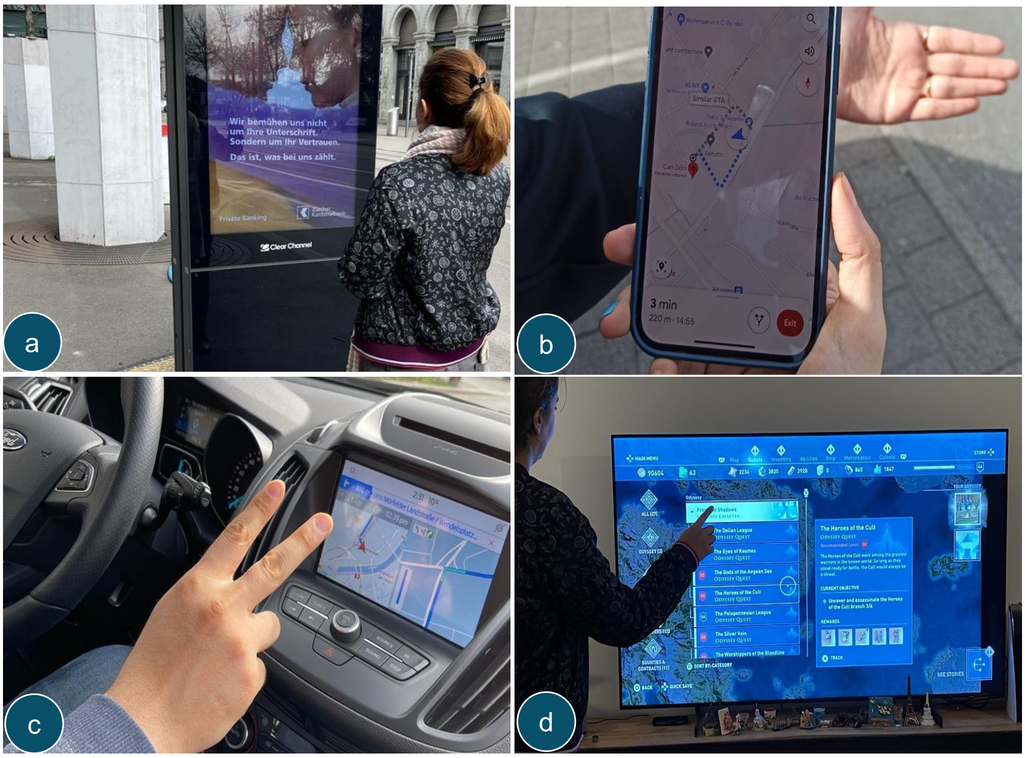



**Figure 1.1:** Gestures enable a variety of interactions performed with the fingers, hands, arms, and the whole body across diverse contexts of use, including mid-air interactions with (a) public kiosk information systems, (b) mobile devices, (c) in-vehicle systems, and (d) smart TVs.

## 1.2 Research Questions and Contributions

Guided by previous considerations, our work aims to enhance the design and generalizability of mid-air gestural interaction systems. The main question guiding this work is as follows.

**How can we design and evaluate a generalizable vocabulary of mid-air gestures that supports intuitive, consistent, and transferable interaction across diverse pervasive computing environments?**

Specifically, our work considers different perspectives, which we break down into the following research questions.

**RQ1:** To what extent are mid-air gesture sets consistent across application domains, and how can this inform the design of a consensus gesture set transferable across domains?

**RQ2:** How do users' preferences and motivations for gesture-based interactions in Smart homes differ depending on the scenarios?

**RQ3:** How are mid-air gestures physically performed across users and repetitions, and what do these performance characteristics reveal about their suitability for designing robust gesture-based interaction systems?

With **RQ1**, we examine the consistency of mid-air gesture sets across different application domains in human-computer interaction, aiming to understand the potential for establishing a consensus gesture vocabulary. We investigate this through two approaches. First, we systematically reviewed 172 papers that reported mid-air gesture sets and classified 1,728 gestures collected from these papers using a taxonomy that capture key characteristics such as form, flow, and nature. From this analysis, we identified gestures that appear repeatedly across domains and assessed their consistency using agreement rates. We further verified the potential for a consistent and transferable gesture set through a scenario-based elicitation study in a smart home laboratory, where participants performed gestures for everyday activities across diverse scenarios. Together, these approaches provide a structured basis for comparing gesture vocabularies, uncovering patterns of overlap, and investigating the potential for a unified and transferable gesture set for broader applications.

With **RQ2**, we examine how users' preferences and motivations for gesture-based interactions differ across scenarios within the smart home use case. Specifically, we investigate how situational contexts, such as cooking or exercising, shape user attitudes toward mid-air gestures, including the devices they prefer to control, the commands they most often elicit, and the benefits they perceive (e.g., hygiene, comfort, or minimizing interruptions). To this aim, we conducted a scenario-based gesture elicitation study with 20 participants in a fully equipped smart home laboratory, covering four distinct daily-life contexts: cooking, fitness training, TV and entertainment, and home office work. Based on this study, we analyze how contextual needs influence gesture choice and perceived value.

With **RQ3**, we look towards gesture performance, i.e., the physical execution of mid-air gestures and how motion characteristics vary across individuals and repetitions. Specifically, we investigate how users perform commonly proposed mid-air gestures in interactive systems in terms of aspects such as trajectory, timing, articulation, and spatial extent. To this aim, we present a novel mid-air gesture dataset titled MAGIC: Multimodal Air Gesture Interaction Corpus, a high-fidelity multimodal dataset developed by us, consisting of 4,265 mid-air gesture instances collected from 42 participants using synchronized electromagnetic tracking and depth cameras. By creating this dataset, this thesis analyzes intrapersonal and interpersonal variability in gesture execution to inform the design of robust gesture interaction systems applicable across diverse contexts.

## 1.3 Methods and Approach

In order to complete the research goals, our work first surveyed existing gesture vocabularies, conducted scenario-based elicitation studies and semi-structured interviews to capture user requirements, developed a multimodal motion capture setup, organized lab studies to evaluate gesture performance, and applied qualitative and quantitative data analysis methods.

### Literature Survey and Synthesis

At the outset of the research, we conducted a structured analysis of existing gesture vocabularies employed across various domains, including automotive systems, gaming platforms, medical applications, and others. This allowed us to identify patterns, consistencies, and challenges in gesture design and informed the selection of candidate gestures with potential for cross-domain applicability.

### Semi-Structured Interviews and Elicitation Studies

We conducted a scenario-based elicitation study supported by semi-structured interviews. Participants engaged with four simulated smart home contexts—cooking, fitness training, entertainment, and home office—within a fully equipped smart home laboratory. They were interviewed to capture reflections on their behavioral routines, interaction preferences, and motivations for using mid-air gestures. This approach enabled us to explore gesture design in context and to ground our findings in users' lived experiences and expectations.

### Sensor-Based Motion Capture and Dataset Creation

To support empirical analysis of gesture execution, we developed a multimodal motion capture setup combining 14-sensor electromagnetic tracking with Multiple depth cameras. This configuration enabled the high-resolution recording of full-body posture and 3D hand trajectories during mid-air gesture performance.

### Data Analysis

Our data analysis approach combined qualitative and quantitative methods grounded in established practices from HCI, gesture recognition, and participatory design. Drawing from literature review insights, user study data, and motion-captured recordings, we employed a different data evaluation strategy. To identify consistent gesture patterns, we used structured coding and agreement measures. Thematic analysis was applied to interpret qualitative interview data across elicitation studies. For the motion-captured dataset, we extracted kinematic features and conducted time-series comparisons to assess gesture performance. We further evaluated gesture recognition using LSTM-based models in both user-dependent and user-independent settings to assess recognition performance.

Statistical testing, guided by data distribution properties, was used to examine differences across participants, gesture types, and study conditions. For the user studies, users' subjective experiences were assessed through post-experiment questionnaires [VD18], along with custom demographic questionnaires, which were analyzed to evaluate the potential influence of background factors on the analysis outcomes.

## 1.4 Publications

The following peer-reviewed publications constitute the main part of this thesis.

- **Hosseini, M.**, Ihmels, T., Chen, Z., Koelle, M., Müller, H., & Boll, S. (2023). Towards a consensus gesture set: A survey of mid-air gestures in HCI for maximized agreement across domains. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–24. `https://doi.org/10.1145/3544548.3581420`
- **Hosseini, M.**, Müller, H., & Boll, S. (2024). Controlling the rooms: How people prefer using gestures to control their smart homes. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18. `https://doi.org/10.1145/3613904.3642687`
- **Hosseini, M.**, Valkov, D., Degraen, D., Müller, H., Koelle, M., & Boll, S. (2025). MAGIC: A dataset capturing mid-air gesture performance for interaction and feature analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT).*`https://doi.org/10.1145/3831634`

## 1.5 Thesis Outline

This dissertation is divided into seven chapters, as illustrated in Figure 1.2. We conclude Chapter 1 with this introductory chapter, where we outlined the motivation, problem statement, research questions, and methodology guiding this dissertation.

Chapter 2 provides the background and related work that situates this dissertation. It reviews the historical development of mid-air interaction, prior work on gesture elicitation, existing taxonomies and classification frameworks, as well as motion capture techniques and motion gesture datasets. Together, these foundations laid the research directions and methods pursued later in the dissertation.

Chapter 3 presents a systematic literature review on gesture transferability and cross-domain applicability and, following established methodologies (e.g., PRISMA), addresses our first research question (RQ1). Here, we define the scope and corpus, present our data analysis framework, and synthesize findings into a consensus-based gesture set.

Chapter 4 addresses our second research question (RQ2) on user preferences for integrating mid-air gestures into everyday smart home scenarios. We present the details of our scenario-based elicitation study and explore which devices and commands are most suited for mid-air interaction, how legacy bias influences gesture proposals, and how ecological validity and scenario familiarity shape user behavior. It also considers gesture consistency across scenarios in the smart home, therefore contributing to Research Question 1 (RQ1).

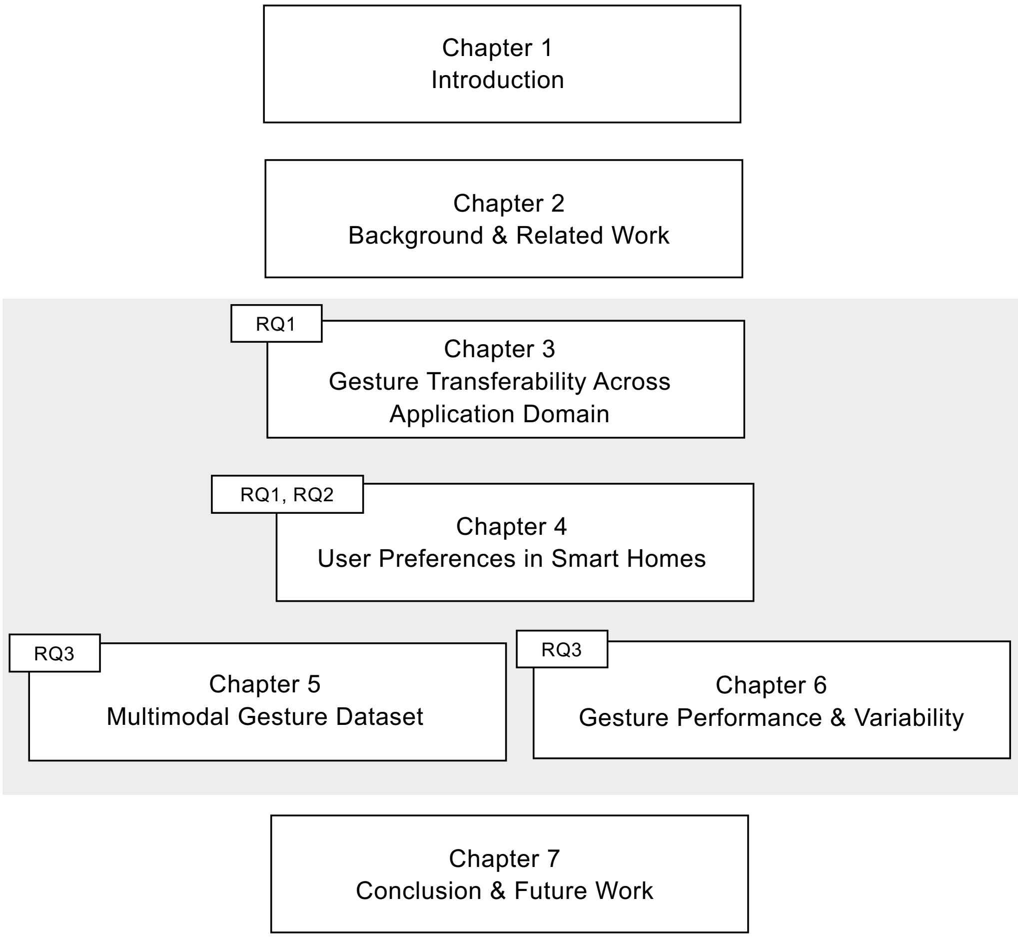



**Figure 1.2:** Thesis structure and corresponding chapters for investigating each research question.

Chapters 5 and Chapters 6 of this dissertation address our third research question (RQ3) concerning the performance and evaluation of mid-air gestures regarding motion features. Chapter 5 presents the creation of a multimodal dataset of mid-air gestures, designed to provide a high-fidelity, open-access resource for analyzing gesture properties in our work and in future studies. Chapter 6 builds on this dataset to investigate gesture performance by applying a range of kinematic features to characterize movement properties and to examine both inter- and intra-person variability in gesture execution, thereby informing their suitability for different contexts (e.g., public vs. private systems).

Finally, Chapter 7 provides an overview of the key conclusions drawn in this thesis and suggests potential areas for future research.

# 2 Background and Related Work

This chapter reviews prior research relevant to the design, capture, and analysis of mid-air gestures in HCI. Specifically, it discusses four areas that form the foundation for this thesis's contributions. The first area traces the historical development of mid-air interaction, from early conceptual visions to contemporary sensing systems (Section 2.1). The second examines approaches to designing mid-air gestures, including both early expert-driven techniques and more recent user-defined gesture elicitation methods (Section 2.2). The third area reviews sensing technologies for gesture capture, based on contact-based, vision-based, and wireless approaches, and highlighting their relative strengths and limitations (Section 2.3). The fourth and final area focuses on gesture datasets captured using different sensing technologies, summarizing existing resources and techniques for evaluating gesture performance and variability (Section 2.4). Each section concludes with observations on how the reviewed work informs and motivates the research presented in this thesis.

## 2.1 Historical Development of Mid-Air Interaction

The concept of mid-air interaction originated in the early days of virtual reality (VR) research. In a paper entitled "The Ultimate Display" [S+65], Sutherland refers to notions such as "in many cases, the computer program needs to know which part of a picture the person is pointing at," the "kinaesthetic display," and the idea that "the ultimate display would, of course, be a room within which the computer can control the existence of matter." Building on these early ideas, the Media Room of the MIT Architecture Machine Group [Bol80] pioneered multimodal interaction by integrating speech recognition with position sensing in the "Put That There" demonstration. In 1986, Vincent John Vincent and Francis MacDougall [Ges24], researchers at the University of Waterloo in Canada, demonstrated the first applications of vision-based, accessory-free tracking of the human body. By 1991, they had gradually developed a whole new genre of performance technology, including various virtual performance instruments such as musical scales, soundscapes, virtual drum kits, virtual juggling, and virtual dance experiences [Pro12]. In 1997, the concept of Tangible Bits was introduced by Professor Hiroshi Ishii, which integrated digital information with physical objects and environments [IU97]. It enables users to interact directly with digital content by manipulating everyday objects and architectural surfaces. These ideas later evolved into G-Speak, an early spatial operating environment that enabled mid-air gesture interaction, invented and developed by John Underkoffler, a former student of Ishii and illustrated in the science fiction film Minority Report [IU97].

The first commercial application of mid-air interaction was the Nintendo Wii gaming platform, released in 2006 [JT12]. It comes with a remote controller (the Wiimote), which tracks the player's hand. The Wiimote detects the player's physical movements using built-in motion sensors and infrared tracking. These movements are translated into corresponding actions of an on-screen avatar, which are displayed on the television during gameplay, such as in sports simulations. By enabling full-body and gesture-based interaction, the Nintendo Wii shifted electronic gaming from the computer desktop to the living room. It addressed diverse user groups, including Children, older adults, and families.

The first affordable, wearable, mid-aid interaction infrastructure was presented at TED in 2009. The Sixth Sense project [MM09] involves a computer-based, wearable, gestural interface that augments the physical world with digital information on demand. This interface can be manipulated with hand gestures.

The Microsoft Xbox 360 was originally released in 2005, and in 2010, Microsoft introduced the Kinect sensor as an accessory for the console. The sensor includes a color camera, a depth camera, and a microphone array for capturing localized sound. Later versions of the Kinect sensor include Kinect for Windows (version 1.8), which was released in February 2012, and Kinect for Xbox One (version 2.0), which was released in 2013. The Kinect SDK enables full-body skeleton tracking of 25 points for two users within a 54-degree scene, at a distance of approximately four meters from the sensor [Zha12]. It enables developers to create games and other applications that allow for mid-air interactions.

The Leap Motion depth camera and SDK, which supports real-time motion tracking of users' hands, palms, and fingers, was released in 2012. The Leap Motion is an affordable sensor that is usually placed in front of the user on a desktop computer. In 2016, a new software development kit (SDK) was released that enables the integration of Leap Motion with virtual reality (VR) head-mounted displays (HMDs), such as the Oculus Rift. This enables controller-free, mid-air interactions in VR [WVM+16].

More recently, mid-air interaction has been applied in domains such as automotive user interfaces, for example, gesture control system in the BMW 7 Series, and in smart television systems. Although Microsoft discontinued Kinect production in 2017, alternative sensors such as Orbbec Astra and Intel RealSense have continued to support research and development in gesture recognition.

## 2.2 Designing and Evaluating Gesture Sets for Interactive Systems

This section reviews prior work on gesture vocabulary design, addressing approaches to defining gesture sets and methods for their analysis, including agreement measures, taxonomies, and cross-domain challenges.

### 2.2.1 Approaches to Gesture Vocabulary Design in HCI

A major challenge in designing computer systems that rely on gesture-based interaction is identifying a set of gestures that 1) the system can recognize them with high accuracy and minimal processing time, and 2) users must be able to perform them easily, with low physical effort and strong memorability of the movements involved. Over time, various methods have been used to define gesture vocabularies in different contexts. Early gesture vocabularies in HCI were largely shaped through top-down, expert-driven approaches, in which designers and engineers defined gestures according to system capabilities, ergonomic principles, and symbolic metaphors such as *Put That There* [Bol80], which combined speech and pointing gestures within a predefined spatial interface. Similarly, *Charade* [BBL93] employed glove-based freehand gestures that were manually mapped to remote control functions. In the *DigitalDesk* system [Wel93], touch and hand gestures were crafted to manipulate digital documents projected onto a physical desk, drawing on physical metaphors such as tapping or dragging. Although some exploratory studies involved user observation, direct user participation in defining gestures was relatively rare.

Although these expert-driven methods were valuable for early explorations of gesture-based interaction, they were not necessarily reflective of user behavior and often resulted in unintuitive, difficult-to-learn, or culturally biased gestures. As gesture-based interaction expanded to broader user groups and more diverse contexts, these limitations prompted a shift toward more participatory, bottom-up approaches. The most notable of these are gesture elicitation studies [WARM05; WARM05; WMW09], sometimes also referred to as guessability studies. In these studies, participants are presented with system functionality or commands and asked to propose gestures that would control those functions.

Many of the key concepts of gesture elicitation were first introduced by Nielsen et al. [NSMG03] and later formalized by Wobbrock et al. [WARM05; WMW09]. Wobbrock et al. [WARM05] investigated how to maximize gesture guessability with a bottom-up approach, allowing end-users to propose gestures for actions. A few years later, Wobbrock [WMW09] presented the first application of this method to hand gesture input in the context of surface computing.

Therefore, GESs have become popular as an effective and valuable tool for designing gestures that align with users' behavior and preferences. Since then, many user-defined gesture sets have been developed across different contexts of use, including:

- **Platforms and devices.** Since the inception of gesture elicitation studies, they have often been designed with a focus on particular platforms or devices. For example, Morris et al. [MWW10] reported users' preferences for multi-touch input on interactive tabletops. Vatavu [Vat12] and Zaiţi et al. [ZPV15] investigated mid-air gestures for controlling TV sets. Ruiz et al. [RLL11] examined motion gestures designed for smartphone interaction, and Chen et al. [CCC16] studied gesture vocabularies for large tabletop displays.

- **Environments.** Gestures are typically elicited within specific physical or psychological environments, which can shape the resulting vocabularies. For example, public settings, where social acceptance and audience played an important role [KAC+18]; in-vehicle user interfaces, where creation of ergonomic and easy-to-use interfaces without compromising safety is a challenge [KSJ+20]; or high-risk environments such as operating rooms, where the key challenge is to meet sterility requirements and to create precise user interfaces [SLL+14].

- **User groups.** GES are designed either for a user-independent mode, where no particular participant profile is targeted, or in a user-dependent mode, where particular profiles of users are included. designing gestures for children [CKLP13], older adults [CTCG17], people with motor or visual impairments [KJWL09], and investigations of how cultural background influences proposed gestures [MHWC10].

### 2.2.2 Gesture Analysis and Classification

A core component of end-user elicitation is the concept of agreement, which indicates when gestures proposed by participants are, essentially, the same or at least substantially similar. Agreement has been evaluated in various ways. Wobbrock et al. [WARM05] provided an agreement measure to analyze and interpret elicited data. This measure has been widely adopted by other elicitation studies [SBCS+12; VZ14]. Vatavu and Wobbrock [VW15] later refined this measure to more accurately represent findings, although there is debate around the accuracy of the agreement measures and concerns about over-optimistic estimates [Tsa18]. Further measures for the level of agreement are agreement rate [FLW12], max-consensus and consensus-distinct ratios [Mor12], or chance-corrected coefficients of agreement, and the growth rate of the dissimilarity-consensus curve. Cafaro et al. [CLK+14] suggest framing the gestures through “Embodied Allegories” to reach a better consensus among users. A review of a large number of publications by Villarreal-Narvaez et al. [VNVVW20] shows that agreement measures have largely been used for analyzing gesture elicitation data. This dissertation extends the related work by taking a cross-domain perspective and by making use of agreement rates as a tool to investigate the consistency of gestures across domains.

Beyond these quantitative measures of consensus, researchers also rely on gesture taxonomies to qualitatively classify and interpret elicited gestures. There are several high-level classification schemes of gesture types and gestural interactions that are applicable to HCI. Yet, classification can differ depending on the context. A work in this space by Karam and schraefel [K+05] provides a comprehensive classification of gesture types in the field of HCI. Starting from the observation that gestures have different forms in different application domains, they classify gestures based on four different dimensions: gesture style, input, application domain, and output systems. While they categorize gestures based on application domain, they do not investigate consistency and don't explicitly compare across domains. In their seminal work on gesture elicitation, Wobbrock et al. [WMW09] look at a specific application domain, surface computing, and address a gesture classification based on the user's mental model and their behavior. Ruiz et al. [RLL11] addressed gesture classification for motion gestures for mobile interaction, where they addressed physical characteristics of gestures as dimensions of their taxonomy. Bostan et al. [BBC+17] focused on gesture classification of micro-gestures, while Arefin et al. [ASLX+16] defined a taxonomy based on mapping features of gestures and their physical characteristics for smartwatches. Despite the extensive discussion of gesture classification, less focus has been placed on the differences between the proposed gestures in different application domains. Our work adopts dimensions from existing taxonomies appropriate for building our classification system and reports which kinds of gestures are being used to interact with interactive objects across different domains of HCI.

In summary, prior research has advanced the design and analysis of gesture sets, but most efforts remain tied to specific devices, applications, or contexts and do not (necessarily) span different devices, scenarios, or domains. Still, very few studies have been conducted to examine the consistency and transferability of gestures across domains or devices. For example, Vatavu [Vat13a] conducted two guessability studies for 22 common home entertainment commands in two different settings of home entertainment, one using handheld devices and the other using freehand gestures. He found that out of 22 gestures, nine were consistent in both settings. Dingler et al. [DRSH18b] designed touch gestures for basic reading controls that successfully transferred across three different devices: smartphones, smartwatches, and smart glasses. Vogiatzidakis and Koutsabasis [VK19] developed a user-defined gesture vocabulary for controlling smart home devices consisting of 7 devices and 55 commands. They demonstrated that many users had difficulties remembering a large number of commands when in real-life scenarios. As a result, they found that most users employed their mental models when interacting with the ecosystem of devices, resulting in high consistency in their proposed gestures for commands that applied to various devices.

Yet, as computing technology increasingly permeates diverse aspects of everyday life, users will expect gestures to transfer not only across devices but also across environments and domains. Achieving such cross-domain consistency is essential for a seamless user experience, but remains an open challenge that this dissertation directly addresses. To address this gap, this thesis investigates cross-domain and scenario consistency in mid-air gestures in Chapters 3 and 4.

## 2.3 Sensing technologies for capturing gestures

Many sensing technologies have been developed to capture and analyze motion. Contact-based sensors capture information through devices attached to the body or objects, and can be used to track positions, forces, tactile parameters, or muscular activity. Unlike contact-based sensors, vision-based sensing systems use cameras or depth sensors to capture motion information without physical touch. Another group is wireless sensing, which detects motion by analyzing changes in radio signals such as Wi-Fi, radar, or UWB. This section provides a closer examination of the functions and characteristics of each sensing device.

### 2.3.1 Contact-Based Sensing Technologies

*a) Hand data glove:* Hand data gloves are electronic wearable devices that utilize a variety of sensors to capture finger flexion, contact points, and, in some cases, spatial orientation in real time. The current products can detect individual finger bends and estimate the hand's position in three dimensions by integrating inertial measurement units (IMUs) or other embedded sensing elements. Glove-based interfaces are widely applied in hand motion tracking and animation, as they enable multi-degree-of-freedom (DOF) representation for each finger with high accuracy, rapid response, and robust operability. For example, Luzanin et al. [LP14] developed a data glove-based recognition system using a probabilistic neural network for VR applications. Cai et al. [CGWS15] proposed a wireless data glove using Xsens sensors for gesture control. It is proven that data gloves are an effective method for capturing hand movements [WZ15; RIAH15]. However, there are still challenges, such as the need for reliable calibration across different hand sizes and the potential reduction of natural hand flexibility when wearing the glove. Figure 2.1 illustrates some of the popular data gloves available on the market.


**Figure 2.1:** Examples of data gloves available in the market. Images taken from product websites of CyberGlove Systems [Sys25], DGTech [Sol25], SenseGlove [Sen25], and Manus [Man25] (accessed September 2025).

*b) Force sensor:* Force sensors are the base for stable manipulation, especially in multi-fingered robotic hands. They measure the forces applied during contact with objects and provide feedback for control. Key requirements for these sensors in manipulation tasks include robustness, sensitivity, and fast response, and in the case of tactile arrays, high spatial resolution. Several sensing techniques are commonly used [KLCZ12]. Capacitive sensors primarily measure displacement and force based on variations in the distance between the upper and lower electrodes, which are caused by changes in external forces. Piezoresistive sensors detect changes in the electrical resistance of the sensing materials, which are often fabricated on silicon or polymer substrates, and are valued for their simplicity of integration but can be temperature sensitive. The piezoelectric pressure sensors generate an electrical charge proportional to dynamic forces such as impacts or vibrations, but are unsuitable for static or quasi-static force measurement. Strain gauges measure changes in resistance due to material strain (deformation), which can be calibrated to infer applied forces, torques, or loads, and are widely used for their robustness and accuracy. Despite their effectiveness, force sensors still face challenges such as nonlinearity, susceptibility to stress, noise, and temperature variations, as well as the difficulty of embedding them compactly in robotic hands without compromising dexterity.

c) *Surface electromyography (sEMG):* Surface electromyography provides technical support for evaluating muscle movement biofeedback by measuring EMG signals on the skin's surface [XJX$^+$18]. The signals reflect muscle contractions and can be analyzed to interpret a person's movement intentions. By learning these patterns, systems such as prosthetic hands and assistive robots can learn how to respond appropriately, enabling gestures and other actions to be controlled. For example, Al-Timemy et al. [ATBEO13] demonstrated the classification of multiple hand gestures using sEMG channels for prosthetic control. Hu et al. [HLL$^+$16] implemented real-time sEMG-based control of dexterous grasping. Although sEMG provides a direct channel to user intent, its performance is often affected by electrode displacement, signal crosstalk, and skin conditions, which remain challenges for robust, long-term use. Figure 2.2 shows examples of applications of sEMG.

d) *Electromagnetic trackers:* Electromagnetic tracking systems, such as Viper Polhemus [Pol], use a low-frequency magnetic field from a transmitter, and detect it with small sensors attached to the body to measure six degrees of freedom (6DOF) motion [FHB$^+$14]. They provide real-time position and orientation data for any rigid body, and can be used in various applications, including biomechanics, robotics, motion tracking, and surgical navigation. However, their performance is constrained by a short effective range, high susceptibility to distortion from nearby ferromagnetic or conductive materials, and reduced accuracy in electromagnetically cluttered environments.

e) *Optical marker systems:* An optical marker-based motion capture technique has been used to track motion in controlled conditions, utilizing calibrated cameras

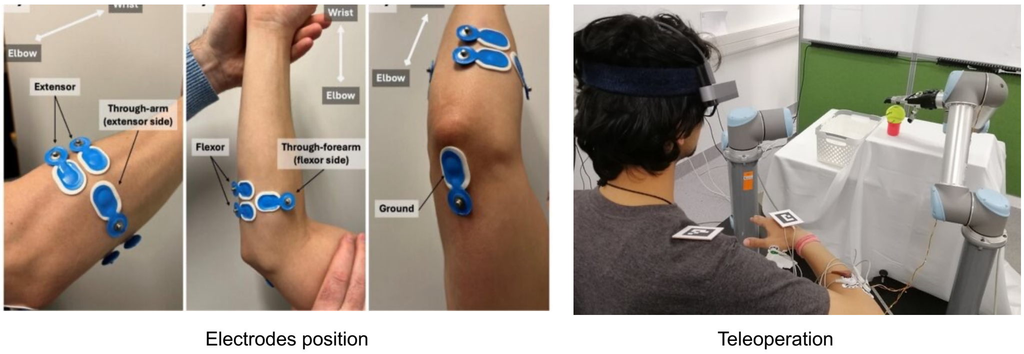



**Figure 2.2:** Examples of Surface Electromyography (sEMG) Electrode Placement and Teleoperation Setup [DGKL19; FRYJ24].

specifically designed to track markers placed on the human body. They are capable of providing reliable, accurate, and fast estimates of joint positions when markers are clearly visible and lighting is well controlled [CPMX07]. For instance, Kuo et al. [KCIO+03] developed a non-invasive tracking system that employed skin sensors and surface markers to capture 3D quantitative motion data. Metcalf et al [MNC+08] introduced a kinematic model that relied on surface marker positioning, applying standard computational methods to determine the marker locations. Optical markers can accurately track human hand motions; however, their use is limited by issues such as errors caused by skin deformation, strict placement requirements, restricted measurement space, the need for specialized cameras, sensitivity to lighting conditions, vulnerability to occlusions, and the inconvenience of wearing markers or suits. Some example images captured with optical markers are shown in Figure 2.3.

**Figure 2.3:** Examples of Hand motion tracking using optical markers in object manipulation tasks [CPMX07].

### 2.3.2 Vision-Based Sensing Technologies

a) *Ordinary cameras:* For motion capture without physical contact, ordinary cameras, primarily RGB and stereo cameras, are widely used. RGB cameras capture three color channels (red, green, blue), typically using independent CCD sensors for each channel. The color information is obtained through variation and superimposition of signals from these three sensors, producing high-resolution video streams that can be analyzed for gesture and motion recognition [JWZ+14]. They are affordable, provide rich color and texture information, and achieve high frame rates. However, human gestures are highly dynamic, and monocular RGB systems often suffer from occlusion and depth ambiguity due to perspective projection. These issues can be solved by employing multiple RGB cameras positioned at different angles to reconstruct 3D human motion.

   A stereo camera consists of two digital cameras of the same specifications. They estimate depth by triangulating disparities between corresponding points in the two images. Stereo systems are compact and portable due to their fixed baseline and internal calibration. However, their performance is limited by the relatively small angle between the two lenses, which reduces accuracy in resolving large occlusions. Additionally, they are sensitive to lighting variations, require careful calibration, and struggle to cover large occlusion areas due to the limited angle between lenses [XJX+18].

b) *Depth cameras:* Compared to ordinary cameras, depth cameras capture not only color but also per-pixel depth information and are much faster and easier to deploy 3D vision systems for human–hand motion (HHM) analysis. Depth sensing is achieved through three main approaches, including time-of-flight (TOF) techniques, structured light, or stereo vision techniques. Unlike monocular or stereo RGB cameras, depth cameras provide a direct depth map aligned with the color image, therefore reducing the need for complex 3D reconstruction algorithms. For example, a ToF-based automatic method was proposed by Oprisescu et al. [ORS12] for recognizing predefined hand gestures. Kinect and Intel RealSense, as the most typical depth cameras, provide synchronized color and depth images and are widely adopted in applications such as computer graphics, human–computer interaction (HCI), robotics, and computer vision research [JJLL15]. Figure 2.4 shows a comparison of RGB and Depth Images Captured with Kinect V1 and Kinect V2. Depth cameras provide the advantage of markerless operation and robust per-pixel depth measurements. However, they remain sensitive to ambient lighting conditions, specular or shiny surfaces, and can suffer from depth noise and occlusion, leading to reduced accuracy in cluttered environments.

c) *Infrared systems:* Infrared-based systems, such as the Leap Motion controller, use infrared LEDs and cameras to capture fine hand and finger movements with sub-millimeter accuracy [WBRF13]. Unlike depth cameras, which rely

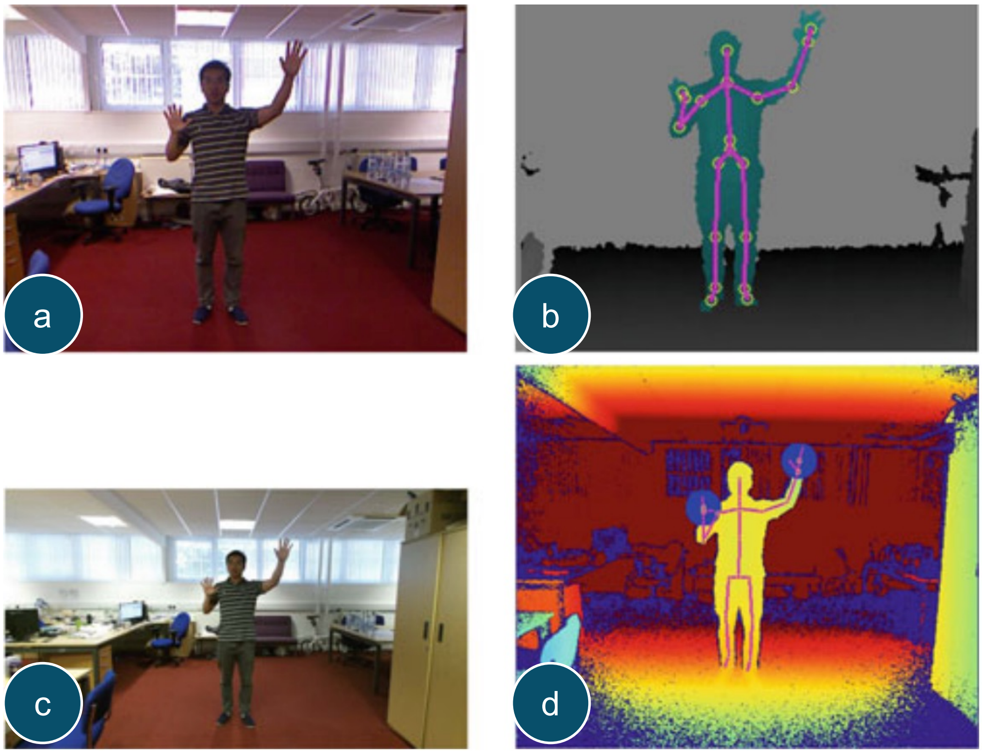



**Figure 2.4:** Comparison of RGB and Depth Images Captured with Kinect V1 and Kinect V2. (a) RGB image (640 × 480 pixels) from Kinect V1. (b) Depth image (320 × 240 pixels) from Kinect V1. (c) RGB image (1920 × 1080 pixels) from Kinect V2. (d) Depth image (512 × 424 pixels) from Kinect V2 [LJJ+17].

on structured light or time-of-flight sensing, Leap Motion uses infrared optics to achieve precise above-surface tracking for near-field interaction. They are compact and affordable, and they enable hand motion tracking with high precision for near-field applications, therefore supporting precise three-dimensional interaction in contrast to conventional multi-touch interfaces, which are restricted to two-dimensional input. Researchers have explored diverse applications of Leap Motion; for example, Chen et al. [CMYF15] implemented a hand gesture–based control system using Leap Motion that enabled the operation of a virtual Universal Robot UR10 through a mathematical processing model. Mapari et al. [MK15] developed an Indian Sign Language recognition framework using Leap Motion to recognise the positional information of the HHMs. However, infrared systems operate within a limited capture volume directly above the sensor and remain sensitive to occlusion (e.g., overlapping fingers), fast movements, and external infrared interference, which can reduce accuracy.

### 2.3.3 Wireless Sensing Technologies

a) *Radar-based sensing:* Radar systems transmit radio waves and analyze reflected signals to infer motion. Techniques such as frequency-modulated continuous-wave (FMCW), Doppler, and millimeter-wave (mmWave) radar have been applied to hand gesture recognition and enable the detection of dynamic gestures and even subtle finger movements [S+80]. Google Soli, for example, demonstrated sub-millimeter tracking using a 60 GHz radar chip [LGK+16]. Radar-based sensing is robust to lighting, works in darkness, and can capture micro-motions that may be challenging for optical systems. These properties make radar suitable for applications such as touchless gesture control in-car human–machine interfaces [WWF+22], health monitoring [AMK+15], and robotic interaction [CDT+22]. For example, Tsang et al. [TCS+21] implemented an FMCW millimeter-wave radar gesture recognition system using spiking neural networks, achieving over 98% accuracy on benchmark datasets. However, radar systems face challenges, including lower spatial resolution compared to vision-based methods, potential interference in cluttered radio environments, high computational requirements for real-time signal processing, and difficulties in distinguishing fine finger articulation with the same precision as camera-based systems.

b) *WiFi CSI sensing:* WiFi Channel State Information (CSI) methods exploit standard WiFi signals to detect motion [WLC+14]. Variations in human gestures and body movements alter the multipath propagation patterns of WiFi signals, and these changes can be analyzed to classify activities without requiring dedicated wearable or vision-based sensors. This makes WiFi sensing attractive for smart homes and pervasive computing, as it leverages existing infrastructure. For example, Wang et al. [WLS+15] demonstrate activity recognition in indoor environments using WiFi CSI. It shows that WiFi sensing can classify daily activities with high accuracy. However, WiFi-based sensing generally provides only coarse spatial resolution, is highly sensitive to environmental interference, and struggles to distinguish overlapping movements. Its accuracy depends strongly on the deployment environment and device placement, which remain challenges for reliable real-world applications.

c) *Ultra-Wideband (UWB) sensing:* Ultra-Wideband (UWB) impulse radars transmit extremely short pulses across a wide frequency spectrum, enabling fine temporal resolution [Gez08]. UWB systems are capable of tracking subtle motions, gestures, and even respiration, and have been widely explored for applications in indoor localization, healthcare monitoring, and human–computer interaction [MGKP15; ASDA+19]. Their ability to provide accurate ranging and penetrate common materials makes them attractive for robust gesture recognition in cluttered environments. However, UWB requires dedicated hardware, typically operates only over short ranges, and regulatory constraints in certain regions may restrict its use.

## 2.4 Mid-air Gesture Datasets

Existing gesture datasets vary in terms of scale, the number of classes, type of annotations, sensors used, and the domain of gestures. Early video-based datasets, such as the Actions Dataset [SLC04] and the Actions as Space-Time Shapes Dataset [BGS+05b], are fundamental resources for studying gesture recognition. These datasets contain simple, predefined full-body actions (e.g., walking, jumping, waving) and have several limitations. For example, they are captured from a single viewpoint, use low-resolution grayscale video, and lack coverage of fine-grained or semantically meaningful hand gestures relevant to contemporary human-computer interaction scenarios.

To address the need for more comprehensive data, new datasets have been developed that incorporate a variety of modalities and capture complex real-world actions and hand gestures. For instance, the Cambridge Hand Gesture Dataset [KC08] consists of 900 image sequences of 9 symbolic hand gesture classes, captured with a fixed RGB camera and defined by three primitive static hand shapes and three primitive motions. The SKIG dataset [LS13] comprises 1080 RGB-D video sequences of 10 gesture classes involving geometric and shape-drawing movements, captured using a Kinect sensor by 6 subjects, and characterized by three hand shapes, three types of background, and two types of arm motion. MSRGesture3D [WLC+12] consists of 336 video sequences of 12 dynamic hand gestures, performed by 10 subjects using a Kinect sensor. The gestures are inspired by American Sign Language (ASL), and include actions like "milk" or "store." The dataset provides depth and skeleton information, but focuses on coarse motion, with no fine finger articulation or semantic gestures like "OK" or "Tap."

Since 2011, the ChaLearn Gesture Challenge has provided a series of large-scale datasets. The ChaLearn LAP IsoGD and ConGD datasets [WZZ+16; MYG+16] provides the most significant number of subjects, gesture classes, and samples, comprising over 47,000 RGB-D sequences and 249 gesture types. However, these datasets were not explicitly designed for human-computer interaction. Instead, they include gestures from various application domains, such as sign language, signals to machinery or vehicle, and pantomimes.

In addition to video-based gesture datasets, several other multimodal datasets have been developed to study finer motion details by incorporating various sensor modalities, such as electromyography (EMG) sensors or inertial measurement units (IMUs). The Ninapro dataset [AGC+14], for example, focuses on hand gestures and was collected using electromyography (EMG) and inertial sensors to capture the fine-grained muscle activities and joint movements associated with both basic wrist/finger motions and complex functional grasps. The HANDS dataset [HGS+20] is a multimodal collection designed to support the development of prosthetic hand control by providing synchronized eye-view and hand-view images, videos, EMG, and IMU data to model and infer human grasp intent. CapgMyo [DLG+21] offers dense sEMG recordings across 128 electrodes and

EMG-IMU-EPN-100+ [VBLVCB22] pairs EMG and IMU data of both static and dynamic hand gestures.

**Table 2.1:** Comparison of public gesture datasets

| Datasets | Samples | Labels | Subjects | Modalities | View | Focus |
|---|---|---|---|---|---|---|
| Actions [SLC04] | 600 | 6 | 25 | RGB(Grayscale) | 1 (frontal) | Full-body actions |
| Actions as Space-Time Shapes [BGS+05b] | 90 | 10 | 9 | RGB(Silhouette) | 1 (frontal) | Full-body actions |
| Cambridge [KC08] | 900 | 9 | 2 | RGB | 1 (frontal) | Symbolic hand gestures |
| SKIG [LS13] | 1,080 | 10 | 6 | RGB-D | 1 (frontal) | Geometric gestures |
| ChaLearn MMGR [EGB+13] | 13,858 | 20 | 27 | RGB-D, Audio | 1 (frontal) | Cultural (Italian) |
| ChaLearn IsoGD/ConGD [WZZ+16] | 47,933 | 249 | 21 | RGB-D | 1 (frontal) | Mixed (signs, pantomimes) |
| Ninapro [AGC+14] | 5,000+ | 52 | 40+ | sEMG, IMU | N/A (non-visual) | Muscle activity |
| HANDS [HGS+20] | 413 | 5 | 11 | RGB, EMG, IMU | 2 (eye-view, hand-view) | Grasp |
| CapgMyo [DLG+21] | 4,240 | 8 | 23 | sEMG (128 channels) | N/A (non-visual) | Static Hand and Finger Motions |
| EMG-IMU-EPN-100+ [VBLVCB22] | 30,600 | 12 | 85 | EMG + IMU | N/A (non-visual) | Static and dynamic hand |
| **MAGIC (ours)** | **4,265** | **18** | **42** | **Depth + EM sensors** | **3 (left, right, back)** | **Static and dynamic hand, frequent in interactive systems** |

These datasets provide rich, fine-grained motion information, but their design is primarily oriented toward functional or biomechanical gestures such as grasping, flexion, or muscle-driven hand postures, which are relevant in biomedical or assistive contexts, but not necessarily representative of the communicative mid-air gestures explored in human-computer interaction studies. While some datasets include individual gestures such as pointing, waving, or a thumbs up, to the best of our knowledge, no publicly available dataset contains all 18 uni-manual gestures examined in this thesis that were identified as the most frequently proposed in gesture studies across diverse HCI domains. Additionally they do not comprehensively capture the execution of these gestures using high-fidelity motion tracking. A comparison of our dataset with some related gesture datasets is provided in Table 2.1.

Furthermore, most work utilizing motion datasets has primarily aimed to improve gesture recognition performance, rather than to understand how gestures are physically executed. Only a few studies have focused on analyzing gesture performance itself. For example, Vatavu [Vat17] examined users' whole-body gestures using spatial, kinematic, and posture-based features. Generally, kinematic and geometric attributes are treated as intermediate features for classification models. However, their potential as analytical constructs for studying gesture consistency variability remains largely underutilized. In this thesis, our analysis focuses on the kinematic properties of gestures to explore consistency, variation, and patterns of natural performance across users and repetitions to assess the generalizability of gestures, identifying those that are user-independent and suitable for robust real-world human-computer interaction systems.

# 3 Gesture Transferability and Cross-Domain Applicability

A large body of research exists investigating gestures for a specific device, context, application, or tracking technology. Those gesture sets are usually designed for and targeted at specific settings. They do not (necessarily) span across different devices, scenarios, or domains. Very few works have tackled *cross-device consistency*. Vatavu [Vat13a] conducted two device-specific guessability studies, one for handheld interaction and one for free-hand interaction in a home entertainment scenario. He found that only 9 out of 22 gestures were consistent in both settings. Dingler et al. designed a set of touch gestures that successfully transferred across smartphones, smart watches, and smart glasses [DRSH18a]. However, as computing technology continues to reach all aspects of people's lives, users will expect gestures to transfer not only between different devices but also between different contexts and environments, i.e., *cross-domain consistency*. Reaching *transferability* of gesture sets is essential for a seamless user experience and will require both *cross-device consistency* and *cross-domain consistency*.

The vision of *universally valid gesture sets* that remain consistent across domains and environments is compelling, particularly in the context of smart environments and pervasive computing technologies, where interfaces might blend in with furniture, appliances, or other everyday objects. Here, cross-domain consistency is especially crucial for mid-air gestures as it is often hard to communicate which gestures are available, e.g., in a specific room. As a result, it is difficult for users to discover, learn, and memorize new gestures. For instance, a user controlling an ensemble of lights in an intelligent living room would use one gesture to select the lights in question, then a gesture to indicate the desired functionality (e.g., light intensity), and then a gesture to set the intensity level (e.g., the degree of brightness). Ideally, the same sequence of gestures would yield exactly the same result when performed in any other room, even including other types of rooms, such as hotel or hospital rooms, offices, classrooms, gymnasiums, or large lecture halls. Extending this line of thought, functionalities would transfer across applications: the user might use the same gesture to modulate other intensity levels, such as the volume of music while driving in the car or watching TV. Despite this, it remains unclear how far we are from implementing this vision. It remains unknown whether and to what extent existing gesture sets presented in HCI literature already possess the potential to transfer across domains and what challenges will need to be overcome to achieve full cross-domain consistency.

To address this, this chapter presents a systematic review of 172 papers on mid-air gestural interaction to synthesize existing gesture sets, characterize their properties, and evaluate their potential for cross-domain consistency. We begin by constructing a representative corpus of mid-air gesture studies and extracting their documented gesture sets, Section 3.1.1. We then describe the data analysis and synthesis approach, including the classification of gestures using an adapted taxonomy and the calculation of agreement rates to evaluate consistency of gestures both in cross-domain settings and within individual application domains, Section 3.1.2. Based on this analysis, we identify a set of 22 gestures with high cross-domain agreement, which we propose as a candidate consensus gesture set in Section 3.2. Finally, we discuss the implications of these findings for gesture set design, highlighting challenges and opportunities for achieving cross-domain consistency in future systems in Section 3.3.

Parts of this chapter have been published as a full paper in the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems by Hosseini et al. [HIC+23]. It is worth noting that this topic is also explored in Chapter 4 in Sections 4.2.4, 4.2.5, and 4.2.6.

## 3.1 Review Methodology

This section outlines the methodological approach used to conduct the review. We describe how relevant studies were identified and selected, and how the gesture corpus derived from these studies was coded in terms of its characteristics, and how agreement rates were analyzed to evaluate cross-domain consistency.

### 3.1.1 Paper Selection Process

We employed the PRISMA guidelines' five-phase process [PMB+21] to create our corpus: *(i) exploration*, *(ii) identification*, *(iii) screening*, *(iv) eligibility*, and *(v) inclusion*.

#### Exploration

We exploratively surveyed the existing research landscape on mid-air gestures to identify prominent keywords and publication outlets. We closely examined the search results of different combinations of "mid-air" and "gestural interaction" as well as synonyms in combination with domain-specific keywords such as "smart home" or "mixed reality". This process ensured that our search terms (cf., Table 3.1) would be consistent with terms used across different relevant domains. For the final appraisal of search terms, we prioritized seminal publications and frequently cited papers, but we also used our own expertise of gestural interaction as a domain to identify relevant and frequently employed keywords.

We compiled a search term that returned a comprehensive list of 2078 initial search results across all databases.

### Identification

We compiled a search term (cf., Table 3.1) across seven online publication outlets: ACM Digital Library[1], IEEEXplore[2], Taylor & Francis[3], SpringerLink[4], Elsevier ScienceDirect[5], SagePub[6] and MDPI[7]. We targeted all fields, including keywords, title, abstract, and full text of the paper, in our search. We did not employ any restrictions in terms of publication date, but we limited our search to publications available as open access or via institutional subscriptions.

Our search on April 6, 2021, yielded an initial set of 2078 papers across all outlets. After removing broken links and duplicates 2057 papers were kept for screening.

**Table 3.1:** Search terms used to search seven common online publication outlets.

```
("Mid-Air" OR "In-Air" OR "3D" OR "Kinaesthetic") AND ("Gesture
Interaction" OR "Gesture Interactions" OR "Gestural Interaction"
OR "Gestural Interactions" OR "Gesture Elicitation").
```

### Screening

We evaluated the relevance of all 2057 papers that resulted from our search term based on their titles and abstracts. Three of the authors reviewed all references independently and in parallel used the following *screening criteria* for inclusion and exclusion: (1) We included only papers written in English; (2) We included only full papers published in peer-reviewed journals, conferences, symposiums or workshops. Notably, this excluded technical reports, dissertations or other theses, work-in-progress reports, patents, posters, and workshop or tutorial proposals; (3) We filtered for topical match, which at this stage included all papers that (a) employed mid-air gestures to control an interactive system or (b) evaluated and/or compared a set of mid-air gestures. We did not apply restrictions regarding specific body parts, methodologies, user groups, or application domains; (4) We excluded papers on mid-air gestures assisted by a physical medium such as pens, wands, or remote controls (e.g., using the Wii Remote asan input device [BCLJ09; SPHB08]); (5) We excluded papers not pertaining to our domains of interest, human-computer interaction (HCI) or human-robot interaction (HRI), e.g., works on co-verbal gesturing, dancing, or sign-language. We made use of Ouzzani et al.'s Rayyan[8] to facilitate collaborative screening. We determined inter-rater agreement between

[1] ACM Digital Library, `https://dl.acm.org`, accessed 23/01/2023
[2] IEEEXplore, `https://ieeexplore.ieee.org`, accessed 23/01/2023
[3] Taylor & Francis, `https://www.tandfonline.com`, accessed 23/01/2023
[4] SpringerLink, `https://link.springer.com`, accessed 23/01/2023
[5] Elsevier ScienceDirect, `https://www.sciencedirect.com`, accessed 23/01/2023
[6] SagePub, `https://journals.sagepub.com/`, accessed 23/01/2023
[7] MDPI, `https://www.mdpi.com/`, accessed 23/01/2023
[8] Rayyan, `https://www.rayyan.ai`, accessed 23/01/2023

the three authors performing the screening based on Fleiss' kappa, which indicated substantial agreement $\kappa = .665$ (95% CI, .642 to .688), $p < .001$. Discrepancies between the reviewers' assessments were then discussed and conflicts resolved. In total, we kept 288 papers.

### Eligibility

To determine eligibility for further analysis, we reviewed all 288 candidate papers at the full-text level. Again, this was done by three of the authors independently and in parallel. At this stage, we used the following (additional) *eligibility criteria* for inclusion and exclusion: we only included papers: (1) that contain at least a minimal, textual description or documentation of the employed gestures, for example the work by Cheng et al. [CGX$^+$15] presented a gesture-based interface for human-drone interaction without providing information about the gestures employed, and papers (2) that contain a mapping between action and gesture, i.e., indicate which action is executed by a particular mid-air gesture, for example, Datcu et al. [DL13] presented four hand-based mid-air gestures for interaction with mobile augmented reality applications, but did not specify the respective actions. We furthermore excluded papers (3) that gestures are not designed as command, for instance, Xiao et al.'s work [XYT13] investigating natural, non-verbal body language (e.g., reflecting emotional states) in the context of interactions between humans and virtual humans, and papers (4) that were re-used from previous work. Regarding the last case, there were papers by the same authors that, based on their previous study, presented a similar system in the same domain in their follow-up paper. In this case, we retained the primary work, where the gesture set was first presented, and we excluded the latter one.

In addition, we re-evaluated all full texts based on our *screening criteria* (see above), which excluded works of the wrong publication types or works making use of a physical medium (e.g., wands). After careful review, this resulted in a corpus of 172 papers. Again, we determined Fleiss' kappa, which indicated substantial agreement between the three reviewers: $\kappa = .662$ (95% CI, .602 to .721), p $< .001$. Any discrepancies were discussed and resolved.

### Inclusion

Our final corpus consists of N=172 papers containing a total of 1728 gestures and corresponding commands (see Section 3.1.2, Gesture Coding, for details). The papers span a time period from 1996 (Sharma et al.'s 3D display for molecular biologists [SHP$^+$96]) to 2020 (15 papers in total). Table 3.2 shows the list of mid-air gesture studies among different journals and conferences. A large majority of the papers (92 %) have been published in or later than 2010, which coincides with the first releases of affordable (gesture and motion-sensing) depth cameras such as the Microsoft Kinect [CRS17]. The publications are split over 100 venues. Conference full papers (151, 87%), with ACM CHI (10 papers), IEEE 3DUI, IEEE

VR, ACM AVI, and ACM DIS (5 each), are the largest contributors to the corpus. Journal articles (e.g., IEEE Access, IMWUT) account for 12% of the corpus (21 papers). The data extraction was performed by three researchers independently and in parallel (each paper by one reviewer).

**Table 3.2:** Publication venues of selected papers.

| **Conferences:** | **N=151** | **87.79%** |
|---|---|---|
| CHI (10 papers per venue) | 10 | 5.81% |
| 3DUI, VR, AVI, DIS (5 papers per venue) | 4*5=20 | 11.63% |
| ISS, ICMI, UIST (4 papers per venue) | 3*4=12 | 6.98% |
| PerDis, OzCHI, TVX, UbiComp, VRCAI, SMC, SUI (3 papers per venue) | 7*3= 21 | 12.21% |
| AutomotiveUI, CHItaly, ICIMCS, NordiCHI, APCHI, AH, MUM, IUI, ITS, ROMAN, ICPR, MobileHCI (2 papers per venue) | 12*2=24 | 13.95% |
| HCII, DAPI, INTERACT, IEOM, IHCI, ISCID, ICCIT, CW, AICT, ICARCV, ICCSN, URAI, VL/HCC, ICHMS, CHIuXiD, MM, AINAW, CISIS, ICMCS, PETRA, EurolTV, CompSysTech, CSNT, 3DTV, DigitalHeritage, DASC, Mobility, Expresive, IVCNZ, SACI, ICIM, ACE, JCSSE, Cyberworlds, ISCC, ISCMI, ROBIO, ICSIPA, IHC, SVR, ICCSNT, RCAR, IECON, Chinese CHI, TEI, ISMARW, CGS, BioVis, Muc, IndiaHCI, MIG, HAI, Mindtrek, BMEiCON, CENIM, EMBC, W4A, DIVAnet, ISETC, SISY, CVPRW, CCNC, VAST, FG (1 paper per venue) | 64*1=64 | 37.21% |
| **Journals:** | **N=21** | **12.21%** |
| IEEE Access (3 papers per venue) | 3 | 1.74% |
| TVCG, EICS, IMWUT (2 papers per venue) | 3*2=6 | 3.49% |
| Interacting with Computers, Informatics, Systems Science & Control Engineering, Procedia Technology, Procedia Manufacturing, Procedia Cirp, Energy Procedia, VRIH, Procedia-Social and Behavioral sciences, Biomedical Informatics, TACCESS, TOMM (1 paper per venue) | 12*1=12 | 6.98% |

### 3.1.2 Data analysis and synthesis

We applied an adapted classification framework together with agreement rate analysis to evaluate gesture vocabularies extracted from the reviewed papers across domains, which enabled us to identify patterns, consistencies, and factors influencing cross-domain transferability.

#### Gesture Types and Classifications

Characterizing gestures based on different dimensions of a taxonomy allows for better distinguishing between different gestures as well as for describing major attributes of each gesture. After reviewing the literature and collecting gestures, we decided that by incorporating some of the dimensions of Wobbrock [WMW09] and Vafaei [Vaf13] taxonomies and by adapting a few of the dimensions, we could form the basis of a taxonomy for our work. From Wobbrock's taxonomy, we selected BINDING (original format) and FLOW (original format). From Vafaei's taxonomy, we chose TEMPORAL (original format, similar to Wobbrock's Flow), NATURE (original format, modified version of Wobbrock taxonomy), FORM (modified, as a mix of Wobbrock and Vafaei), and BODY PART (modified). There are multiple subcategories within each dimension, which we explain in the following section.

NATURE of the gesture denotes the relationship between gesture and meaning/object and divides into five categories: *manipulative*, *pantomimic*, *symbolic*, *abstract*, and *pointing*. A gesture is *manipulative* when there is a tight and direct relationship between the movements of the gesture and its impact on the object/entity, such as moving two hands away to scale a virtual object. *Pantomimic* gestures mimic a meaningful action or object, such as a hand in a grabbing posture (grabbing a knob) and then rotating it to increase or decrease volume. *Symbolic* gestures visually depict symbols, such as the thumbs-up gesture, the OK sign. *Pointing* gestures are considered as *pointing* gestures that indicate real, implied, or imaginary persons, objects, or directions. For example, it can be used for identifying appliances in ubiquitous computing [VK19]. *Abstract* gestures map gestures to the interactive task arbitrarily. Indeed, there is no manipulative, symbolic, or pantomimic connection between gesture and task, such as making a fist to turn on the TV.

The BINDING dimension is considered the location where the gesture is performed. In *object-centric* gestures, the location is defined with respect to object features. This category mostly covers transformation tasks, such as moving or rotating the objects. In *world-dependent* gestures, the location is defined with respect to the UI space (app. environment) features. Opening a horizontal and vertical menu with swiping the hand from right to left and from up to down, respectively, is classified as *world-dependent*. Indeed, the user needs to know the screen coordinates to perform this gesture. *World-independent* gestures can occur anywhere, such as snap fingers to turn on TV. Gestures that occur in multiple spaces (both *object-centric* and *world-dependent*) are categorized as *mixed* dependencies, such as grabbing an object and moving it to a specific space.

FORM categorizes gestures based on whether and how the position and movement of the body part vary within a gesture. This dimension is defined as uni-manual, which means, in the case of two-handed gestures, that it applies separately to each hand [WMW09]. FORM includes *static pose* gestures, *static pose and path* gestures, *dynamic pose*, *dynamic pose and path* gestures, *stroke* gestures and *multiple*. *Static pose* gestures are performed in only one location, where the pose is held over time, such as pointing with the index finger. *Static pose and path* gestures hold the pose of the gesture while moving the body part, such as swipe hand. *Dynamic pose* gestures change the position of a body part during the gesture, while gestures are performed in only one location, such as closing a fist.

*Dynamic pose and path* gestures consist of changing both position and location, such as closing both hands into fists and then moving them apart. *Stroke* gestures consist of taps/flicks, such as pressing an imaginary button in mid-air. With bi-manual gestures, each hand performs a different form, such as clenching the right hand into a fist and moving the index finger of the left hand.

The FLOW dimension indicates whether the command occurs during a movement or at the end of a movement. In this dimension, a gesture is discrete if the task occurs after the gesture ends, e.g., drawing a circle to turn on the TV. Continuous gestures are gestures that complete the task during the gesture. For example, gestures such as move hands up or down to make commands such as increase or decrease volume are considered to be continuous gestures.

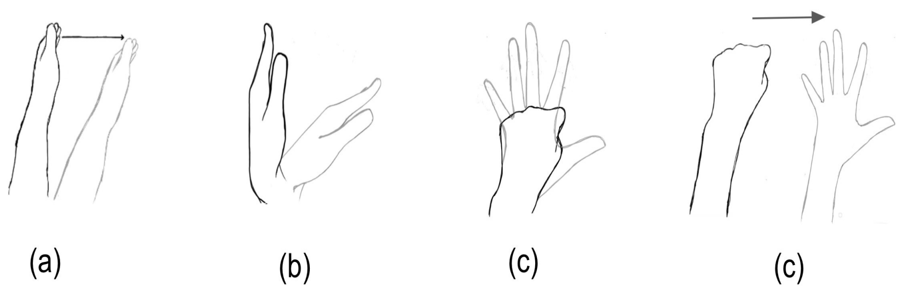



**Figure 3.1:** Classification according to different subcategories of BODY PARTS: (a) The forearm moves to the right (forearm gesture), (b) wave wrist from inside to outside (wrist gesture), (c) Spread the fingers outwards (hand gesture), (d) Spread the fingers outward and move the forearm to the right (hand + forearm gesture)

The BODY PARTS dimension refers to which body parts are involved in a gesture. It distinguishes between hand, wrist, forearm, arm, shoulder, foot, leg, torso, and head. In this thesis, we use the following description for body parts. Hand refers to changing the position of fingers and palm, for example, making a fist. Wrist refers to the functions of the wrist joint, including bending, straightening, moving laterally, and rotating. Forearm refers to the movement of the forearm relative to the elbow; for example, the forearm moves upward and close to the arm, resulting in a decreased angle between them, or it moves downward, increasing the angle relative to the arm. Arm refers to the movement or rotation of the arm relative to the shoulder; for example, moving the arm toward or away from the body. Shoulder refers to shoulder movement relative to the neck; for example, elevating the shoulders to convey I don't know something. Foot refers to movements upward or downward and rotation of the foot at the ankle. Leg refers to the movement or rotation of the leg relative to the hip or knee, including moving the leg toward or away from the median plane. Torso refers to the vertebral column movements relative to the sacrum and hip bone, for example, bending forward. Head refers to turning the head relative to the trunk, including turning the head laterally or turning it back to look straight ahead, or moving the head and neck forward and downward, like looking down.

In our classification, we distinguish between gestures that include one body part and those that include variants of other body parts. For example, pinch with thumb and index finger (hand gesture) vs. move the hand back and forth while pinching the thumb and index finger (hand plus forearm). Figure 3.1 illustrates some of the subcategories of Body Parts: (a) forearm movement, (b) wrist gesture, (c) hand gesture, and (d) hand plus forearm.

#### Agreement Between Gestures

About 65% of the gestures in this work were designed by end-user elicitation studies. In these studies, the functions are shown to users, who are then asked to produce gestures for each action that would execute the function [WMW09]. A key element in end-user gesture elicitation studies is the concept of "agreement", meaning that gestures defined by participants are the same or essentially similar [AMW18]. The level of agreement between participants was the most frequent metric used in gesture elicitation studies. The agreement rate measures the level of consensus between participants for each task. It was introduced by Wobbrock [WARM05] and later improved by Vatavu and Wobbrock [VW15].

To determine the consensus gesture set, this chapter makes use of literature as a basis for calculating agreement rates, not end-user elicitation. After providing the gestures for each referent, we calculated the agreement rate for each referent based on the definition proposed by Vatavu and Wobbrock [VW15]:

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} (\frac{|P_i|}{|P|})^2 - \frac{1}{|P|-1} \tag{3.1}$$

In Equation (1), P is the set of all gestures proposed for referent r. Pi are subsets of gestures that are unique from P. The range of agreement rate varies between 0 (every suggested gesture is different for referent r) and 1 (all suggested gestures are similar for referent r). According to Vatavu and Wobbrock [VW15], the agreement rate level is to be interpreted as low agreement (AR(r)<0.100), medium agreement (0.100<AR(r)<0.300), high agreement (0.300<AR(r)<0.500), and very high agreement (AR(r)>0.500).

## 3.2 Results

### 3.2.1 Participants Overview

We found that most papers reviewed (66 papers) had between 10 and 20 participants in their user studies. Thirty-two papers included fewer than 10 participants. Twenty-one papers included between 20 and 30 participants. Papers that had significantly larger numbers of participants had 127, 74, and 79 participants.

Among the participants, the number of males was about 1.5 times higher than the number of females. About 33% of the participants were students or a combination of students and academic staff. Around 40% of participants were adults with different backgrounds (sports clubs, engineers, designers...). Participants in 4 articles were related to people with special needs (blind people, individuals with developmental disorders), and 30 papers did not report sufficient information about their participants. We found that 28.32% of papers (49 out of 172) did not conduct a user study or did not report the number of participants included in the study. There were 10 papers that conducted 2 user studies.

### 3.2.2 Application Domains Overview

Our structured literature analysis revealed a diverse range of application domains of mid-air gestures. We classified each paper based on the subject or motivating use case into one of 12 high-level application domains. Some terminologies used for application domains are motivated by ACM index terms, which we list in Table 3.3. Here, the classification of application domains was done iteratively by two researchers individually and in group discussions.

The largest cluster of applications focuses on the selection and manipulation of 2D/3D objects and navigation within different representation types (30 papers). A considerable number of papers in this category were systems that primarily focus on mid-air gestures in a 3D space and immersive VR scenarios, such as interaction with a virtual environment displayed on a mobile device [LRP+19], above stereoscopic interactive tables [SLL+14], or virtual assembly systems [LL13]. Some articles examined interaction with a 2D display or projection [CPB+16; SVAG14]; some examined enabling users to navigate into 3D virtual buildings, museums or pseudo-universes [SDNP19; GPI+15]; and some focused on interaction with augmented reality [PVTMV18]. Typical tasks (but not limited to) within this domain include select [KSP20], rotate [PVTMV18], scale [CPB+16], and translate [PCBC13]. One trait of this category was that some systems here were multi-modal [WGO20].

Other application domains that frequently make use of mid-air gestures are large (public) displays (16 papers). These include tabletop displays or interactive walls specifically designed to be placed in public (or semi-public) locations, including museums [GSR+16], exhibitions [CCC16], and various city settings [ACTK15; SKWL17]. These domains require designing intuitive gestures for heterogeneous users with diverse technical backgrounds.

We identified the use of gestures to control entertainment and multimedia systems, including TV sets [VZ14; ZZ19; Vat12] and music playback systems [HLB+10; LNJ17], as prevalent application domains (19 papers). In these works, the entertainment system is the central element at which all interactions are directed – as found in 'classical' living room arrangements. Thus, this application domain also included studies performed in virtual home entertainment scenarios [LNJ17]. Here, we also grouped related activities performed in home entertainment systems, such as photo-browsing tasks performed on a TV while sitting on a couch or in an armchair [FVM12], as well as drawing applications [ASW+18]. In contrast, smart rooms and appliances (10 papers) include a setting in which mid-air gestures are used to control spatially distributed, multi-device environments that do not follow a traditional layout with one appliance at the center. These include smart home appliances [HZLT20; VK19] and household or building devices [AJS13], including lighting control [MBS+11]. Again, we included both real and virtual environments with otherwise comparable characteristics [YLZ+12; SAUK+19].

Applications in the domain human–robot interaction (13 papers) employ gestures to control robotic devices over a distance, such as drones [SMBK18; OKK+16; JILC17; CEZL15], movable displays [SL18], industrial robots or robot arms [TAM+16; AFT+18; ZLS+17; XCL+12; FAAF15], and tele-operation systems [KKR+14; FAAF15].

We distinguished mobile interaction (11 papers) as a separate domain. These include works that focus on mobile interactions, that look into gestures for controlling personal mobile devices, e.g., around-device gestures for smartphones [SSP+14; KRG+12], or wearable gadgets, e.g., smart watches [ASLX+16; KLK+16]. Gestures designed to interact with another person's device (here: smart glasses [KAC+18]) were included in this domain. In contrast, we grouped under other domains the use of wearables (e.g., smart watches [KPWL16]) as proxies to interact with stationary devices or screens (e.g., public displays).

A specific application domain that traditionally is set in a desktop environment and involves manipulation and navigation tasks is 3D modeling (7 papers). We decided to keep this application domain separate, as 3D modeling activities are often performed by expert users. Works contained in our corpus include gestural commands for three-dimensional shape modeling, ranging from CAD applications to industrial and graphics design. In this application domain we see a focus on tasks such as manipulating faces, edges, vertices, and polygons by means of translation, rotation and scaling, as well as other formations, e.g., [CKS16]. This domain focuses on effective, high-fidelity, precise manipulations [LFZG15]. Following the same rationale, we included gaming (5 papers) as a separate application domain. Gamers often develop great expertise in playing. Works grouped here include the use of gestures to control video games [CNVV15] or games set in Virtual or Augmented Reality [AQSS16]. They typically seek a set of Natural Engagement hand poses during game play [OGTCS20], and gesture sets can help to achieve better immersion in the VR-based games [AQSS16].

E-Learning and education (8 papers), medical technology (8 papers), accessibility & assistive technologies (5 papers), and in-vehicle & automotive (11 papers) are other popular application domains where gesture sets are designed for specific (expert) user groups. Education applications have primarily focused on gestural interaction with learning environments, including interactive exploration of electronic learning resources and educational media [LCD+15]. This domain includes: educational computer games to develop frameworks reducing the difficulty for educational computer game developers [ZZLW12]; smart teaching interfaces [FFC19] that design inputs for manipulating lecture materials on a projected display; and interactive applications for learning and understanding 3D geometry and exploring molecular structures [CL19; PKS15; NSTVCV16; SHP+96; SSTO13]. In these applications, the focus most of the time is on pupils and teachers manipulating educational content [CL19; FFC19].

In the medical domain, mid-air gestures have been explored for their promise of supporting touch-less interaction, which is beneficial in clinical environments. Our corpus contains gesture sets developed for various medical devices [SLL+14] and imaging applications [LKH+13] to be used by experts, i.e., doctors and nurses. Similarly, gesture-based interaction techniques have been explored to enable various assistive, adaptive, and rehabilitative applications. Examples include information browsing designed for vision-impaired users [GDMB18] and gesture sets designed to meet the requirements of people experiencing chronic pain [PJ19]. In these cases, users of assistive devices are typically experts in using their devices.

Gesture sets designed for in-vehicle or automotive user interfaces target primary driving tasks [DFGS19], i.e., moment-to-moment control of the vehicle, and secondary non-driving tasks, such as control of in-vehicle infotainment systems (IVIS) [YMG+20]. They are designed to be performed by the driver [YMG+20; MGW17]. Gesture sets designed for this domain highlight the need for (within-domain) consistency: a gesture that controls music in one car model would ideally not start up the hazard lights when performed in another car.

Other papers (29 papers) include those not associated with a specific application domain or those covering general-purpose scenarios associated with multiple application domains.

**Table 3.3:** Application domains of mid-air gestures.

| | |
|---|---|
| **Manipulation/Navigation (30)** | **Large (Public) Displays (16)** |
| – 2D Display: [SVAG14; CPB+16; FCXW09]<br>– VR/3D: [GFS19; SGH+12; OTK+19; AKK+19; MFA+14; ED15; LSS11; LL13; GS19; OGT+17; Khu15; LFJW98; MO13; BKS20; FPDH15; KSP20; LRP+19]<br>– AR: [PCBC13; PVTMV18; CH18; WGO20; SEC+19]<br>– Building/Museum: [SDNP19; LCK+13; PPR+13; GPI+15; CGLA20] | – Projected Display: [NHON14; CFM+12; KF04]<br>– Application in Public Setting: [SKWL17; CCC16; GSR+16; AWB+14]<br>– Curved Displays: [BW10]<br>– Others: [ACTK15; BAEI16; AKM+20; LZL20; KPWL16; PTWL17; CB15; RM17] |
| **Media & Entertainment (19)** | **Smart Rooms & Appliances (10)** |
| – TV: [VZ14; ZZ19; Vat12; PLR16; DSSR16; XDM+19; IIGI11; DSS+19]<br>– Audio/Video Interaction: [HLB+10; LNJ17; WYC+17; RRVL+14; MGDS14; DM15; HHB16; LCZW11]<br>– Photo-Browsing Tasks: [FVM12; KD16]<br>– Drawing Applications: [ASW+18] | – Controlling Multi-Device in Smart Kitchen and Home: [HZLT20; VK19; GVV18; BRS19; AJS13]<br>– Specific Class of Devices or Single-Task Scenarios: [DKK14; GV13; MBS+11]<br>– Navigation in the Virtual House: [YLZ+12]<br>– Interact with IoT Devices: [SAUK+19] |
| **Human-Robot Interaction (13)** | **Mobile Interaction (11)** |
| – Human-Drone Interaction: [SMBK18; OKK+16; JILC17; CEZL15]<br>– Interaction with Industrial Robots or Robot Arms for Robot Navigation: [TAM+16; AFT+18; ZLS+17; XCL+12]<br>– Robotic Teleoperation: [KKR+14; FAAF15]<br>– Controlling a Group of UAVs: [PH17]<br>– Solving Pick-and-Place Tasks: [QTGJ15]<br>– Controlling TV Display Direction: [SL18] | – Typical Tasks in Smartphones: [BMR+12; FGS+20; SSP+14; AHJ+12; LHY+20; WCB11]<br>– Around Mobile Device Interaction: [SSP+14; KRG+12]<br>– Smartwatches: [ASLX+16; KLK+16]<br>– Smart Glasses: [KAC+18] |
| **eLearning & Education (8)** | **In-Vehicle and Automotive (11)** |
| – Educational Computer Game Frameworks: [ZZLW12]<br>– Smart Teaching Interfaces: [FFC19]<br>– Interactive Applications for Learning and Understanding 3D Geometry: [CL19; PKS15; NSTVCV16; SHP+96; SSTO13]<br>– Digital Storytelling Systems: [LCD+15] | – Primary Driving Tasks: [DGS20; DFGS19; HZKB17]<br>– Secondary Non-Driving Related Tasks: [YMG+20; KSJ+20; MGW17; RSES11; GRSW15; HZKB17; SBD+15; WB19] |
| **Medical Technology (8)** | **Gaming (5)** |
| – 3D Medical Image Exploration and Navigation: [LKH+13; RBM13; LFPP+17; SPHK08; TTSI+13; HH18; SLW+14]<br>– Operating Table Control Systems: [SLL+14] | – Video Games: [FPT12; CNVV15]<br>– VR Based Games: [AQSS16; OGTCS20; SKWA18] |

Continued on next page

**Table 3.3 – continued from previous page**

| **Accessibility & Assistive (5)** | **3D Modeling (7)** |
|---|---|
| – Elderly People: [Lee18]<br>– Individuals with Impairments or Disorders: [SVA+18; PJ19; GDMB18; FSAP11] | – 3D Architectural Urban Planning: [SW11; LFZG15]<br>– Others: [NFUK99; NNK16; CKS16; WLK+14; DVL16] |
| **Others or Not Specified (29)** | |
| – Mix: [CXH15; BBC+17; CLH+20; BKM09; SK98; KS99; DFA15]<br>– Data Analysis/Exploration: [LCG19; WFSN19; CSS+16]<br>– Shape Display Interaction: [LLD+11]<br>– Understanding Specific Body Part: [YYYS18]<br>– Others: [RWP07; RA12; JKK+20; LXS+13; MBGFY19; CMPT12; GPY+17; AHH+18; KHI+12; YWF+16; RMFK14; SAMK11; KKD14; IWSB15; CT17; RGDA14; FNGA12b; DAE+08] | |

### 3.2.3 Agreement per Referent

We extracted a total number of 1728 gestures and corresponding commands from the papers in our corpus, omitting gestures that received low or no consensus in gesture elicitation studies.

Various terms are used in the literature for interaction (including task, command, affect, etc). To be consistent with those terms in our work, we grouped those with similar concepts into one category and called them "referent" in this chapter. For example, for "increase volume", "brightness up", and "heater value up", "increase" was assigned as a referent. Two researchers analyzed all the commands collected from the literature, then agreed on a referent for each command. The referents were derived entirely bottom up from the data.

Afterward, one researcher grouped the referents into 6 categories based on their concept, including Transformation, Navigation, Confirmation, Selection, Adjust Settings, and Actions. We took the Selection and Transformation categories from [PCBC13] and extended them with Navigation, Confirmation, Adjust Settings, and Actions. The grouping was done so that referents at the same level were not compared or overlapped. For example, if a gesture such as tapping results in acceptance, it will be classified as accept/confirm and not selection, or if we decompose selection, it will not be compatible with any referent on navigation, such as rewind. To avoid bias and divergent interpretations, researchers constantly consulted each other when coding and categorizing data.

The set consists of more than 100 referents, Table 3.4 includes those that exist at least in 3 papers.

Table 3.4 lists these referents and their agreement rate. The first two columns identify each referent and its related command category. Column 3 presents the definition of the referent that specifies what tasks are included. Column 4 gives a sample gesture for each referent. Column 5 lists the number of papers for each referent, and the last column shows the agreement rate for each referent. We calculated the agreement rate for each referent based on the definition described in Section 3.1.2.

Gestures were then clustered to identify identical groups of gestures within each referent separately to calculate the agreement rate (AR). We carefully considered how prior works classified gestures for calculating agreement. Consequently, to identify similar gestures, instead of simply coding the gestures found in the literature we used the concept of a "gesture theme." We used "gesture theme" for grouping the gestures [PVTMV18] proposed based on the mental model in previous papers. To the best of our knowledge, there is no comprehensive taxonomy to specify different mental models of gestures. We clustered the gestures using a bottom-up, inductive analysis approach to identify their themes based on a mental model. As gestures sometimes vary, even if only in a very minor way, following a strict coding of these gestures is not useful. This is especially true in our case, where gestures belong to an extensive and varied range of application domains derived from a multitude of publications.

A gesture theme, on the other hand, is a particular mental model or idea of one or multiple gestures, i.e., gestures placed within the same theme can be distinct in how they are executed, but their mental model and main idea are similar. Mental models are sometimes affected by the referent, so each referent is grouped separately. For instance, one of the themes we identified for Zoom was Pull/Squish. Pull/Squish was performed in four different ways: move open hands to or away from each other, regroup or open all fingers, pinch in or out with fingers, or grab two hands together and then move them away from each other. One other theme for Zoom we identified was move close/away (make a grabbing gesture with a hand and move it closer to or away from the eyes; move the head away from or closer to the screen). As a result of our coding gestures to themes, gestures within a particular mental model differ mostly in terms of body part (hand, leg, etc.), path direction (straight vs. flexible), path flow (multiple strokes vs. continuous), hand posture (pointing one finger, spreading or closing fingers, etc.), and movements (fixed sign vs. drawing).

To cluster the gestures, first we conducted an initial coding round where two researchers carefully evaluated one-third of all gestures independently and chose the appropriate theme for each of them. We then held a code adjustment session where two coders discussed with each other and made a mutual decision when they had a conflict. Afterward, one of the researchers coded the rest of the gestures to assign the themes.

The process of obtaining agreement rates is described in Section 3.1.2. As a reminder, according to Vatavu and Wobbrock [VW15], the agreement rate level can be interpreted as low agreement ($AR(r)<0.100$), medium agreement ($0.100<AR(r)<0.300$), high agreement ($0.300<AR(r)<0.500$), and very high agreement ($AR(r)>0.500$).

Table 3.4 shows that among different referents, 11 (19.64%) were with a very high agreement rate, 11 (19.64%) were with a high agreement rate, 23 (41.07%) were with a medium agreement rate, and 11 (19.64%) were with a low agreement rate. Dichotomous tasks in most cases, such as next/previous, increase/decrease, group/ungroup, and accept/reject, had almost the same level of consensus. All of the tasks in the transformation category and adjust setting had a high or very high AR. In the confirmation category there were no high AR at all.

**Table 3.4:** An overview of referents.

| Category | Referent | Definition | Example | # | AR |
|---|---|---|---|---|---|
| **Transform** | Translate | Movement or the dragging of an entity | Push one hand forward [WGO20] | 59 | 0.45 |
| | Rotate | Rotation of entities | Rotate grab hand | 56 | 0.54 |
| | Scale | Decrease or increase the size of object | Close Up two hands [CPB+16] | 27 | 0.78 |
| | Move Cursor | Cursor movement on the screen | Index finger move around [LFZG15] | 12 | 0.6 |
| | Minimize | Reduce the size of object to smallest amount | Close hand [SW11] | 4 | 0.4 |
| | Maximize | Expand object to its full size | Move thumb and index finger from closed to open [XDM+19] | 4 | 0.3 |
| **Navigate** | Pan/scroll | Navigate within a single view | Palm move [AWB+14] | 31 | 0.50 |
| | Zoom | Changes the scale level of an entity | Two hands move apart [DFA15] | 39 | 0.49 |
| | Next | Go to next item, channel, web page,... | Hand swipe right [CXH15] | 43 | 0.52 |
| | Previous | Go back to previous item, channel, web page,... | Hand swipe left [CXH15] | 41 | 0.49 |
| | Fast forward | Move forwards through the tape/video faster that usual | Point right with thumb [DSS+19] | 7 | 0.14 |
| | Rewind | Wind a tape/video back to the beginning | Point left with thumb [DSS+19] | 8 | 0.15 |
| | Main menu | Returning to the main menu | Swipe from right to left [CCC16] | 4 | 0.19 |
| | Home Screen | Going to home screen | Making a fist and open it [FGS+20] | 6 | 0.05 |
| | View control | Going to home screen | Pinch and move around | 10 | 0.22 |
| | Locomotion | Moving from place A to B (robot, avatar, VR) | put left leg in front of the right leg [AQSS16] | 18 | 0.27 |
| | Accept/Confirm | Agreeing in binary queries | Thumb up [PCBC13] | 20 | 0.14 |

*Continued on next page*

**Table 3.4 – continued from previous page**

| Category | Referent | Definition | Example | # | AR |
|---|---|---|---|---|---|
| **Confirm** | Reject | Answering “no” in binary queries | X symbol [DSSR16] | 11 | 0.15 |
| **Select** | Single select | Select a single entity or location | Closing hand into a fist [CNVV15] | 77 | 0.22 |
| | Multiple select | Select group of entities | Tap one after another [LNJ17] | 4 | 0.21 |
| | Deselect | Cancel the selection | Opening hand [LFPP+17] | 16 | 0.37 |
| **Adjust Settings** | Increase | Rise in amount of volume, light, temperature,... | Move hand up [HZLT20] | 42 | 0.33 |
| | Decrease | Reduce in amount of volume, light, temperature,... | Move hand down [HZLT20] | 41 | 0.42 |
| **Action** | Turn on | Turn on any device | Snap fingers [HZLT20] | 18 | 0.07 |
| | Turn off | Turn off any device | Rub hands to turn off heater [GVV18] | 16 | 0.17 |
| | Play | Playing multimedia content | Hand tap [YMG+20] | 23 | 0.12 |
| | Stop | Stopping temporary or permanent an action | Halt gesture (show palm) [VK19] | 39 | 0.12 |
| | Phone Call | Answer an incoming call | Phone pose [KHI+12] | 12 | 0.14 |
| | Delete | Remove or erase entities | Throwing away [RGDA14] | 11 | 0.17 |
| | Undo | Reverse the previous action | Wave hand [SW11] | 9 | 0.14 |
| | Ungroup | Separate objects from each other | Two hands move apart [NFUK99] | 3 | 1 |
| | Group | Move objects together | Two hands move together [NFUK99] | 3 | 1 |
| | Mute | Not giving out sound or speech | Shsss gesture [VK19] | 10 | 0.09 |
| | Unmute | Make a sound or speech again | Opening palm [DSS+19] | 4 | 0.17 |
| | Click (cursor) | Mouse simulation | Rotate wrist [KPWL16] | 10 | 0.13 |
| | Create | Create object or interaction space | Draw outline [WLK+14] | 7 | 1 |

*Continued on next page*

**Table 3.4 – continued from previous page**

| Category | Referent | Definition | Example | # | AR |
|---|---|---|---|---|---|
| | Skip | Skipping directly to a desired entity | Swipe/Slide [BKM09] | 3 | 1 |
| | Activate | Ready to work on object/device | Show number 1 to active interface [DKK14] | 10 | 0.05 |
| | Copy & Paste/ Duplicate | Exact copy of another thing | Hold original object and pull a second hand out [PVTMV18] | 11 | 0.04 |
| | Close menu | Close different type of menu | Swipe [GRSW15] | 8 | 0.42 |
| | Close object | Close document, tools,... | X symbol [XDM+19] | 5 | 0.2 |
| | Open menu | Open different type of menu | Swipe [LNJ17] | 16 | 0.2 |
| | Open object | Open document, tools,... | Grab and open [DAE+08] | 5 | 0.06 |
| | Open door/window | - | Swipe/Slide [AJS13] | 5 | 0.47 |
| | Cut A | Removes the selected data from its original position | Snap index and middle finger [PCBC13] | 6 | 1 |
| | Cut B | Slashing an object | Knife gesture [SKWA18] | 5 | 0.42 |
| | Start | Begin navigating | Point forward [DFGS19] | 4 | 0.1 |
| | Take picture | Take screenshot, selfie, image of object | Forming a square with both hands [SMBK18] | 6 | 0.33 |
| | End call | Reject, Hang-up or end current Call | Flick to the left [GVV18] | 7 | 0.09 |
| | Switch | Switch between multiple functions or state | Select one of the fingers [BBC+17] | 8 | 0.04 |
| | Toggle | Switch between binary functions or state | Downward bounce [CXH15] | 4 | 0 |
| | Info-more | Get more information on an entity or location | Hold (briefly) one arm up [ACTK15] | 4 | 0.17 |
| | Help | Ask (invoke) system for help | Draw question mark [Vat12] | 3 | 0 |

*Continued on next page*

**Table 3.4 – continued from previous page**

| Category | Referent | Definition | Example | # | AR |
|---|---|---|---|---|---|
| | Exit | Exit from system or application | X symbol [LZL20] | 3 | 0 |
| | Deactivate | Unable to respond | Ok gesture to deactivate virtual guide [SDNP19] | 5 | 0 |
| | Trigger | Invoke or summon an object | Tap index and middle to summon the spinbox [GPY+17] | 7 | 0 |

### 3.2.4 Cross-Domain Consensus Gesture Set

We constructed a consensus gesture set of gestures collected from the literature. To construct the gesture vocabulary, we selected referents with a high or very high AR (containing 22 referents). For each of these referents, we selected gesture themes with a frequency of more than 25%.

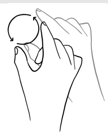
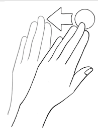
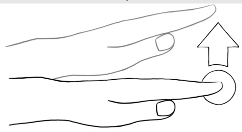
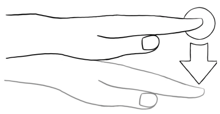
Table 3.5 summarizes the 22 referents. We listed each referent with AR>0.3 and selected the most frequent gesture themes. In those cases where a theme had multiple variations (in terms of execution, body part, uni-manual or bi-manual, and direction), we noted them in the variation part of the table. For the various gestures introduced for a referent, we determined the percentage of agreement with the introduced theme. We also determined the nature of each gesture theme based on our taxonomy. A gesture theme can differ in other dimensions of taxonomy depending on its variation, so we have omitted adding other dimensions of taxonomy. For example, the main idea of "scaling" is Pull apart/Squeeze, which can be done in different ways. One way, for example, is to use one hand pinching fingers in or out (dynamic pose). Another is to use open hands moving away from or close to each other (static pose and path), or by first grabbing both hands together and then moving them away from each other (dynamic pose and path).

Since the goal is to achieve cross-domain consistency, users may not be able to perform bi-manual gestures in many applications or may not prefer gestures made with other parts of the body. Table 3.5 shows uni-manual gestures in most cases, and the bi-manual gestures are only for those cases without an alternative uni-manual variation in the reviewed literature. Since we intended to present one of the executions of gesture themes already found in the literature, we did not add an alternative uni-manual gesture based on our speculations in Table 3.5. Some other gestures are visually different, but the theme seems identical. For example, Next and Open Door are both slide/swipe, and Zoom and Scale are both pull apart/squeeze.

The nature of most gestural themes in Table 3.5 is manipulative, i.e., gesture movements act on the objects themselves. Example themes are *group, ungroup, translate, pan/scroll, scale, rotate, move cursor*, and *zoom*. After that, Symbolic gestures were the most common. Examples here are *skip, next, previous, create, minimize, maximize, increase*, and *decrease*. Pantomimic gestures can be seen in themes for *cut*. Abstract gestures and pointing gestures were not among these themes.

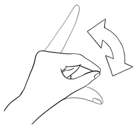
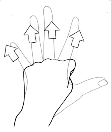
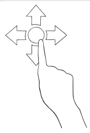
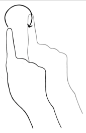
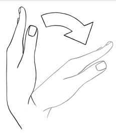
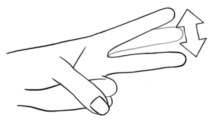
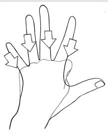
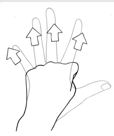
To understand how this 22-gesture set is mapped to the different referents across different domains, we identified the 3 most frequent gestures for each referent and determined their distribution among different domains. Appendix A gives the 3 most frequent gestures for each referent. A comparison of gesture distribution across different application domains reveals just a few conflicts between the gestures we proposed for a given referent (table 3.5) and the most frequent gestures identified between different application domains. They include *translate* in Accessibility and Adaptive Systems and In-Vehicle and Automotive, *rotate* in Human-Robot Interaction, *increase and decrease* in Manipulation/Navigation, *pan/scroll* in Manipulation/Navigation and Medical Technology, *deselect* in 3D Modeling, and *take picture* in Mobile Interaction.

**Table 3.5:** Consensus gesture set compiled from the agreement rate analysis sorted in descending order of papers for each referent.

| Referent | **Translate (AR: 0.45)** | **Rotate (AR: 0.54)** | **Next (AR: 0.52)** |
|---|---|---|---|
| |  |  |  |
| Theme(s) | *Swipe/Slide.* | *Turning Movement.* | *Slide/Swipe.* |
| Variant(s) | With finger, hand, arm, leg, unimanual, bimanual | With fingers, hands, unimanual, bimanual | With finger, hand, leg, different direction |
| Occurrence | 63.7% of 135 gestures in 59 papers | 71.9% of 121 gestures in 56 papers | 71.83% of 71 gestures in 43 papers |
| Nature | Manipulative | Manipulative | Symbolic |
| Referent | **Increase (AR: 0.33)** | **Decrease (AR: 0.42)** | **Previous (AR: 0.49)** |
| |  |  |  |
| Theme(s) | *Swipe/Slide.* | *Swipe/Slide.* | *Slide/Swipe.* |
| Variant(s) | With finger, hand, arm | With finger, hand, arm | With finger, open hand, leg, different direction |
| Occurrence | 53.42% of 73 gestures in 42 papers | 62.85% of 70 gestures in 41 papers | 69.86% of 73 gestures in 41 papers |
| Nature | Symbolic | Symbolic | Symbolic |
| Referent | **Zoom (AR: 0.49)** | **Scale (AR: 0.78)** | **Pan/Scroll (AR: 0.50)** |

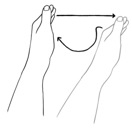
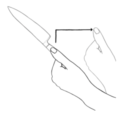
*Continued on next page*

**Table 3.5 – continued from previous page**

| | | | |
|---|---|---|---|
| |  |  |  |
| Theme(s) | *Pull Apart/Squeeze.* | *Pull Apart/Squeeze.* | *Swipe/Slide.* |
| Variant(s) | With fingers, unimanual, bimanual, arm | With fingers, hands, unimanual, bimanual | With finger, hand |
| Occurrence | 67.82% of 87 gestures in 39 papers | 88.31% of 77 gestures in 27 papers | 69.35% of 62 gestures in 31 papers |
| Nature | Manipulative | Manipulative | Manipulative |
| Referent | **Deselect (AR: 0.37)** | **Move Cursor (AR: 0.6)** | **Close Menu (AR: 0.42)** |
| |  |  |  |
| Theme(s) | *Open Hand.* | *Move or Slide Hand Around.* | *Swipe/Slide.* |
| Variant(s) | – | With finger, hand | With finger, hand, arm, unimanual, bimanual |
| Occurrence | 61.9% of 21 gestures in 16 papers | 77.78% of 18 gestures in 12 papers | 66.67% of 12 gestures in 8 papers |
| Nature | Pantomimic | Manipulative | Symbolic |
| Referent | **Create (AR: 1)** | **Open Door (AR: 0.47)** | **Cut A (AR: 1)** |
| |  |  |  |
| Theme(s) | *Draw/Defining Area.* | *Swipe.* | *Snap Index and Middle Finger.* |
| Variant(s) | With finger, hand, arm, head, leg, unimanual, bimanual | – | – |
| Occurrence | 100% of 30 gestures in 7 papers | 66.67% of 6 gestures in 5 papers | 100% of 6 gestures in 6 papers |
| Nature | Symbolic | Symbolic | Pantomimic |
| Referent | **Minimize (AR: 0.4)** | **Maximize (AR: 0.3)** | **Ungroup (AR: 1)** |
| |  |  |  |
| Theme(s) | *Close Hand.* | *Move Apart Fingers.* | *Move Hands Apart.* |

*Continued on next page*

**Table 3.5 – continued from previous page**

| | | | |
|---|---|---|---|
| Variant(s) | – | With thumb and index, all fingers | – |
| Occurrence | 60% of 5 gestures in 4 papers | 60% of 5 gestures in 4 papers | 100% of 3 gestures in 3 papers |
| Nature | Symbolic | Symbolic | Manipulative |
| Referent | **Group (AR: 1)** | **Skip (AR: 1)** | **Cut B (AR: 0.42)** |
| |  |  |  |
| Theme(s) | *Move Hands Together.* | *Swipe/Slide.* | *Slit/Knife.* |
| Variant(s) | – | With finger, arm | – |
| Occurrence | 100% of 3 gestures in 3 papers | 100% of 3 gestures in 3 papers | 66.67% of 9 gestures in 5 papers |
| Nature | Manipulative | Symbolic | Pantomimic |
| Referent | **Take Picture (AR: 0.33)** | | |
| |  | | |
| Theme(s) | *Make Frame with Both Hands.* | | |
| Variant(s) | – | | |
| Occurrence | 57.14% of 7 gestures in 3 papers | | |
| Nature | Symbolic | | |

### 3.2.5 Are Agreement Rates Consistent Across Domains?

We calculated agreement rates in different application domains for all referents listed in Table 3.4 . Figure 3.2 shows a heat map plot of the results. Green (above 0.3) indicates high agreement, red (below 0.29) indicates low agreement. Domains with only one paper per referent (i.e., too few to calculate AR) are marked as "*" (asterisk). Domains with no papers for a referent are marked as "NA".

In all 22 referents with high ARs from our candidate set, 9 had high ARs in each application domain. These were *move cursor, scale, rotate* (from transformation category), *zoom* (from navigation category), *create, close menu, cut, take a picture* (from action category), and *decrease* (from adjust setting). Several referents had a high value in cross-domain as well as in most domains and had a low AR in only one or two domains. For example, *next* and *previous* had low AR in accessibility and adaptive systems and in e-learning and education domains (2 papers, AR=0 for both). *Increase* had medium AR in manipulation/navigation (3 papers, AR=0.17). *Pan* had only medium AR in in-vehicle and automotive (AR=0.24), and *deselect* had low AR in media and entertainment (3 papers, AR=0.048).

Some referents, for example *selection* and *stop/pause*, had a high degree of variability across domains. It seems that for *stop*, AR is lower when we want to stop a device such as a drone [JILC17] or a robot [FAAF15], compared to when we want to stop digital content [LCZW11] such as media or a cursor.

Among the referents with a lower AR across all domains, most had a low AR even in separate application domains. However, there were some exceptions. Across all domains *open menu* had AR=0.2. But in manipulation/navigation (2 papers) and in-vehicle and automotive (3 papers), it was very high (respective AR=1 and AR=0.5). Across all domains *accept/confirmation* had medium AR (AR=0.14), but in media and entertainment (4 papers) and mobile interaction (2 papers) it was high (respective AR=0.3 and AR=0.33). We do not see that in its dichotomous task (i.e. *rejection*); *rejection* still has low AR in media and entertainment (4 papers). Across all domains *close object* had AR=0.2 but in media and entertainment(2 papers) it was high (AR=1). Across all domains *turn on* and *turn off* had low AR (respective AR=0.06 and AR=0.17) but they had high AR in media and entertainment (4 papers *turn on*, 3 papers *turn off*) and in-vehicle (2 papers *turn on*, 2 papers *turn off*). Across all domains *view control* had medium AR but it was high in manipulation/navigation (2 papers); and *main menu* is high in large display interaction (2 papers).

Among all the domains, 3D modeling had the most number of referents with a high rate and the mobile interaction had the least. Here are the percentages of referents with high AR in each domain: 3D modeling: 87.5%; robotic: 85.7%; gaming: 75%; large display: 66.67%; manipulation/navigation: 61.1%; medical technology: 60%; in-vehicle and automotive: 58.8%; media and entertainment: 57.7%; e-learning and education: 57.1%; smart room: 54.55%; accessibility and adaptive systems: 50%; mobile interaction: 50%.

### 3.2.6 Does the Task Affect Agreement Rates?

As can be seen in table 3.6, in some referents we used standardized “meta-meanings” to summarize similar tasks, for example for tasks “increase volume”, “increase lightness”, and “increase speed” we used “increase” as a general referent. To understand how the proposed gestures for a referent differ in different settings (tasks), we calculated AR for different tasks of a referent (i.e., the same referent in different settings). As can be seen in Table 3.6, the tasks in most cases do not have an effect on the agreement rate except in next page/slide and previous page/slide, in which AR was medium and low, respectively.

### 3.2.7 Gesture Types per Domain

We manually classified each gesture based on dimensions of the taxonomy described in Section 3.1.2. Our goal was to identify the characteristics of mid-air gestures in different application domains. For coding the gestures based on taxonomy, three researchers individually coded gesture sets, 15% of them by the first researcher, 15% by the second


**Figure 3.2:** Agreement rate of collected referents in different application domains

one, and 70% by the third one. To avoid bias and divergent interpretations, researchers constantly consulted each other when coding and categorizing data.

Figure 3.3 shows the percentage of gestures in each application domain in the 5 dimensions of the taxonomy.

Each gesture was classified based on Flow, Nature, Form, Binding, and Body Part dimensions. In the Flow dimension, Discrete gestures dominate at least slightly in most application domains but, with only a few exceptions, are relatively balanced overall. Only in manipulation/navigation and eLearning and education do continuous gestures account for over 60% of all gestures. Typical referents of these domains are predominantly continuous (such as scale or rotate). It seems that the domains with more discrete gestures have less AR. For example, accessibility, mobile interaction, and smart rooms compared

**Table 3.6:** Comparison of the Agreement rate of referents with different tasks for each referent.

| **Referent/Task** | **AR** | **Theme** | **Nr. papers** |
|---|---|---|---|
| Increase | 0.34 | Swipe/slide | 42 |
| 1- Volume | 0.33 | Swipe/Slide | 32 |
| 2- Speed | 0.38 | Swipe/Slide | 9 |
| 2- Brightness | 0.48 | Swipe/Slide | 6 |
| Decrease | 0.42 | Swipe/Slide | 43 |
| 1- Volume | 0.37 | Swipe/Slide | 33 |
| 2- Speed | 0.27 | Swipe/Slide | 6 |
| 2- Brightness | 0.53 | Swipe/Slide | 7 |
| Next | 0.53 | Swipe/Slide | 43 |
| 1- Menu/List Item | 0.71 | Swipe/Slide | 26 |
| 2- Channel/Audio/Video | 0.4 | Swipe/slide | 17 |
| 3- Page/Slide | 0.17 | Swipe/Slide | 4 |
| Previous | 0.65 | Swipe/slide | 41 |
| 1- Menu/List Item | 0.67 | Swipe/Slide | 22 |
| 2- Channel/Audio/Video | 0.45 | Swipe/Slide | 17 |
| 3- Page/Slide | 0 | - | 3 |

to other application domains have more discrete gestures, and according to the heat map in section 3.2.5 they have lower AR compared to other application domains.

In the Nature dimension, as seen in Figure 3.3, the gesture nature differs depending on the application domain. In this category, manipulative gestures account for at least 28% in 6 domains. This can be explained by the inherent continuous nature of manipulative gestures, which is particularly suitable for performing various geometric and coordinate transformations, especially, in 3D modeling, manipulation/navigation, or medical technology domains. Pointing and abstract gestures are also rather rare in all domains. Some domains use a significant number of Symbolic gestures, especially in smart rooms and media and entertainment, which include about 50% of gestures. It seems that in domains with more Symbolic gestures and fewer Manipulative gestures, AR is less than in other domains, such as smart rooms, media and entertainment, and accessibility and adaptive systems. Some domains use a significant number of pantomimic gestures, especially in accessibility and adaptive systems and gaming, where they account for about a third or more of all gestures.

In the Form dimension, in most application domains except mobile interaction, gaming and accessibility, and adaptive systems, the most common gestures are static pose and path (at least 40% of all gestures). Gaming had more than 50% static poses, mobile interaction had almost 40% dynamic poses, and accessibility and assistive systems had almost 40% dynamic poses.

In the Binding dimension, world-independent gestures were the most common in all application domains, accounting for at least 50% of all gestures. World-dependent gestures were observed in all application domains except for accessibility and assistive systems. Objective-Centric gestures were observed in all application domains, especially 3D modeling and manipulation/navigation, which include about 30% of the gestures.

Mixed dependencies gestures were observed only in manipulation/navigation, media and entertainment, and mobile interaction, with an occurrence less than 5%.

In the Body Part dimension, hand gestures have a higher percentage compared with other gestures in most application domains, except for 3D modeling, large display interaction, manipulation/navigation, and robotics (less than 30%). In those exceptions, arm or arm-hand gestures cover the highest percentage of gestures. The wrist gestures were most repeated (around 10%) in accessibility and adaptive systems, in-vehicle and automotive, and mobile interaction. Forearm gestures are observed in all domains. Foot, torso, head, hand+wrist, and full body gestures were most rare, observed in less than 1% and in just a few application domains.

## 3.3 Discussion

This literature review confirmed that mid-air gestural interaction is extensive, specifically in the area of interaction with remote displays (TVs, large Public Displays, Medical Image Viewer Monitors, etc.) and Home Appliances, In-Vehicle information systems, and Robot Control. To draw deeper conclusions about the applicability and usefulness of this modality of interaction, more specific studies should be conducted in each field as the application range grows.

Now that pervasive computing and using gestures as an interaction modality have penetrated so many different domains in daily life, users need to be able to transfer interaction knowledge from one application or domain to another. Designer- and user-driven gesture studies have been widely used to develop intuitive gestures for individual devices and applications. However, much less work has investigated their use across domain boundaries.

Conducting studies for different devices and domains individually does not necessarily ensure cross-domain consistency of the resulting gestures. We calculated the consistency of mid-air gestures for different types of referents, looking at consistency both across domains and individually within singular application domains. We have found that: (1) there is cross-domain consistency for some of the referents, especially for *move cursor*, *scale*, *rotate*, *zoom*, *create*, *close menu*, *cut*, *take a picture*, and *decrease*; (2) with one or two exceptions, some referents are consistent in most domains, including the referents *next*, *previous*, *increase*, and *pan*; and (3) other referents showed much more variety and lower agreement, which suggests more flexibility in designing gesture languages in those cases. Cross-domain consistency, therefore, may not be achieved for those referents, and other interaction techniques may suit them better. Those referents include *turn on*, *turn off*, *accept*, *reject*, *mute*, *unmute*, and some other abstract referents, such as *activate* or *deactivate*. While agreement rates for some referents are either too low or too high, such as *exit*, or *group*, the occurrence of those referents in various studies is scarce. Thus, although the consensus gesture sets in table 3.5 are not necessarily consistent in each application domain, we believe they can be used as a basis for transferable gesture sets. We defined this consensus gesture set by selecting the most frequent gesture themes from referents that have high or very high agreement rates. In addition, we identified three of the most frequent gestures in each domain for each referent, giving us an overview of how a consensus set maps to the referent across different domains. In general, our analysis suggests that the selected consensus set remains more or less constant between domains,

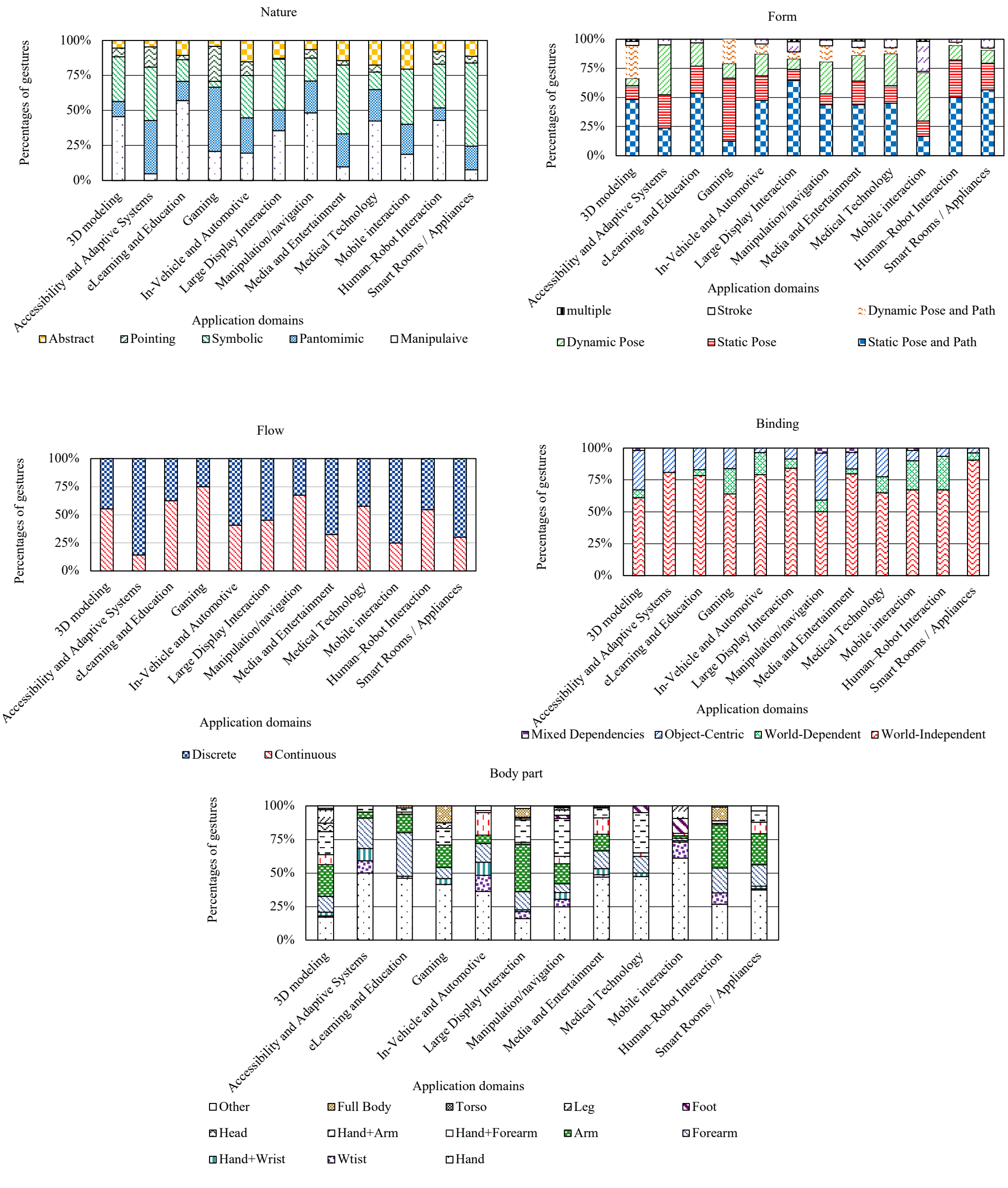



**Figure 3.3:** The percentage of gestures in each application domain in the five-dimensional taxonomy.

because the selected gestures were the most frequent in most domains (or at least the second- or third-most frequent).

### 3.3.1 General Requirement of Using Consensus Gesture Sets

The criteria used to determine the similarity of gestures can affect the agreement rate and consistency of the gestures. Our coding using gesture themes results in a considerably higher agreement rate than the gestures themselves because we count each gesture as unique, regardless of which body part is used (fingers vs. hands vs. fists vs. arms) or if a specific gesture is executed in different ways (draw an X symbol vs. crossing two fingers to represent X). Users may vary their gestures within a specific interaction gesture theme as a response to different device types with different form factors, such as distance from a device, space, and diversity in the nature of application domains. Design implications, therefore, must be considered for gestural interaction systems that do not allow simply reusing a specific gesture. The transferability of gestures across different application domains needs to be considered. For example, in some application domains, such as 3D modeling, gaming, and medical technology, gestures with more freedom of movement are more intuitive and create a more immersive feeling in the user. Designers need to take into account the size of the space/object and the users' proxemics. For example, to rotate virtual objects, users may switch from a single-handed gesture in a small space to a two-handed gesture in a larger space.

In addition, in the consensus set, several gestures are identical but correspond to different referents. Consequently, only a selection of gestures from the consensus gesture set can be used in an application. We did not replace these with the next most repetitive gesture for that referent, because we assumed that not all of the referents in Table 3.5 are going to be used together in the same domain. In the situation we encountered, however, we encourage gesture set designers to choose the gesture from Table 3.5 for the referent with the highest rate of agreement and choose the second most repeated one for the other referents in the Appendix A .

### 3.3.2 Design Guidelines for Gesture Sets

This section is a more in-depth discussion of selected gestures, including our formulation of a set of guidelines indicating possible ways to design mid-air gestures. Each of the following propositions may, of course, benefit from further empirical evaluation to verify their validity.

- **Gestures for Operating Imaginary Objects in Mid-air:** Many of the most frequent gestures were derived from mid-air control of virtual objects. For example, a variety of rotary controllers, slider controls, or buttons are emulated for many referents. They are particularly prevalent in Smart Rooms/Appliance, In-Vehicle and Automotive, Media and Entertainment, and Manipulation/Navigation. Thus, we recommend exploring such gestures along with corresponding virtual objects to create intuitive mappings between gestures and referents.
- **Reuse Touch-Based Interaction Paradigms:** A substantial number of gestures we observed in different domains were variants of taps, directional swipes, slides, and pinch in/out fingers. Mid-air gestures therefore seem to mimic common touch gesture

inputs. Although this is called a legacy bias in the literature [MDD+14], we suggest that mid-air, gesture-based input should take advantage of traditional interaction techniques and touch-screen gestures.

- **Opposing Referents:** Dichotomous referents had almost the same level of agreement, instances of reverse gestures, and similar proportions. This occurred either when the agreement rate was high, such as *zoom in/zoom out*, *scale in/scale out*, *next/previous*, and *increase/decrease*, or when it was low, such as *turn on/turn off*. Designers might therefore want to assign to dichotomous gestures two opposing states (such as closing the fist and spreading the fingers).
- **Similar Themes for Different Referents:** Within the consensus set, similar gesture themes may be assigned to different referents. For example, both Zoom and Scale are squeezed together or pulled apart. Using similar gesture themes with variations for different referents may result in greater memorability of gestures used in various domains.
- **Simple and Common Gestures**: Simple and common gestures such as swipes, slides, points, taps, rotates, pulls, and squeezes were among the gestures most frequently found in different application domains. A good system, therefore, will likely include some variation of this set of well-known gestures.

### 3.3.3 Challenges and Limitations

Through our survey and analysis of literature on cross-domain consistency of mid-air gestures, we identified challenges and open issues for developing a set of transferable gestures. Below are key challenges drawn from 172 papers.

#### Building Interaction Systems in Real Environments

Many systems and elicitation studies in this review were built and tested in controlled lab setups, and they often involved a set of simulation environments [KSJ+20]. Systems and interactions may not transfer and scale well to environments more representative of everyday use, and users may change their preferred gestures outside the lab. Researchers, therefore, need to embed the process of gesture design in the users' typical environments. For example, in a study by Ackad et al. [ACTK15] evaluating mid-air gesture interaction, participants interacted with a system available in a public mall.

#### Conducting User Studies

Many of the papers we examined, especially those focused on designer-made gestures, did not include a user study, and gestures were introduced based on system preference. For example, Dinh et al. proposed gestures based on depth recognition for smart home appliances [DKK14], but did not explicitly consider the users' mental model (i.e., memorability, intuitiveness, discoverability, learnability, and social acceptability). Assessing a user's mental model is essential to design gestural interactions usable across application domains in terms of investigated ambiguity, representativeness, understandability, social acceptability, and comfort of gestures. Designing with the user in the loop will result in better gesture designs than designs optimized purely towards a system's capabilities.

### Reproducing the Adequate Gesture in Context

Some studies in this review were removed from the context of use, thus their real-world application is uncertain [IWSB15; GPY+17]. The number of papers in some domains is low, as well. For example, we identified only 5 papers for gaming and 5 papers for accessibility & assistive systems. Looking at the results section, we see that the referent influences AR. For example, in the transformation category, translate, rotate, and scale had higher AR compared to minimize and maximize. More work is needed in some domains to create a better balance between referents and each application domain.

### Specifying Output Devices

Output devices can vary in multiple ways, including the size of the screen (PC, smartphones, or wall-sized display), 3D scenes in VR/AR, or even real objects in the smart home or in automotive vehicles. Because many papers did not report data related to the output devices they used [RA12], we could not evaluate the effect of different devices on the consistency of the proposed gestures. Going forward, any study should explicitly define the devices used, not only to allow evaluation of consistency across devices, but also to enable researchers to repeat that study with other devices or other user groups.

### Considering a Gesture's Ergonomic Aspects

Another important aspect of gestural interaction design is the ergonomics of the gestures. Research has shown that mid-air gesture systems developed without ergonomic considerations can have a negative effect on the health and experience of users long-term. Indeed, any gesture puts some level of strain on the musculoskeletal system, so that pain, discomfort, and fatigue can result, depending on the duration and amount of strain [Smi96]. The design phase, therefore, should include considerations of the ergonomics of the device or gesture and the anatomy of the body. Table 3.5 is based on the mental model of gestures and lacks the ergonomic aspects. Exciting research has provided guidelines to reduce the risk of disorders stemming from poor ergonomics. For example, using gestures with shorter duration and less use of extended arm range and shoulder movements [HRGMI14] as well as micro gestures [LGK+16] can reduce arm fatigue.

### Limitations and Future Work

Our survey has some limitations. We used the query method to identify the corpus of papers, and the criteria we used for selection necessarily constrained the final corpus. For example, we limited our study to full papers. Some high-quality short papers, therefore, may have been left out. Second, we excluded databases that were not accessible by subscription from our academic institution. Third, the query search was conducted in April 2021, so articles published after that time are not part of this review.

We identified a consensus set of gestures from a number of independent gesture elicitation studies and designers' proposed gestures. This gesture set is derived from literature that reports on different studies using different similarity criteria to cluster the elicited gestures, as well as different sets of referents to elicit those gestures. In the next chapter, we conduct an experimental validation study to investigate the validity of our proposed gesture set.

## 3.4 Conclusion

We conducted a structured literature review of mid-air gestures in HCI and selected 172 related articles. We collected an extensive set of gestures in interactive systems from the literature and clustered them according to the dimensions of taxonomy to understand their commonality across various applications. Using agreement rate evaluations of gestures for more than 50 referents, we constructed a consensus gesture set containing 22 referents. Furthermore, we examined the influence of the application domain on agreement rate and reported challenges in developing or selecting a transferable gesture set based on current studies. The lack of a universal gesture set that can be applied cross-domain is a major challenge in HCI. This chapter and the resulting consensus gesture set contributes to addressing this challenge.

# 4 User Preferences for Mid-Air Gestures in Smart Home Applications

Smart home systems are becoming increasingly popular. They enable us to control household devices and home appliances effectively and intuitively. The current devices can be controlled by using a variety of interaction methods. Wireless remote controls enable us to adjust the temperature or light in our home from a distance. Amazon Alexa and Google/Apple home systems allow users to operate their smart devices through voice command, phone, or interactive displays. In addition to speech and touch user interfaces, previous studies have explored further interaction modalities [RKB+17; MGW17], including mid-air gestural interaction. Figure 4.1 provides an illustrative example of such interaction, showing a user performing mid-air gestures to turn on a kitchen light.

Controlling smart home devices through such mid-air gestures can be beneficial in many cases. For instance, in the case of controlling the TV, by eliminating the need to hold the remote control, which is often not in our hands, and that we must look for, gestures allow for faster and more convenient interaction. During cooking and with dirty hands, mid-air gestures can facilitate touch-free interaction, or they can even simplify user interaction by eliminating the use of buttons. Most of the work in smart homes has been focused on identifying and developing mid-air gestures to control individual devices such as TVs [DDFES15; DSSR16; WW12], music playback devices [LCK+13], or lighting systems.

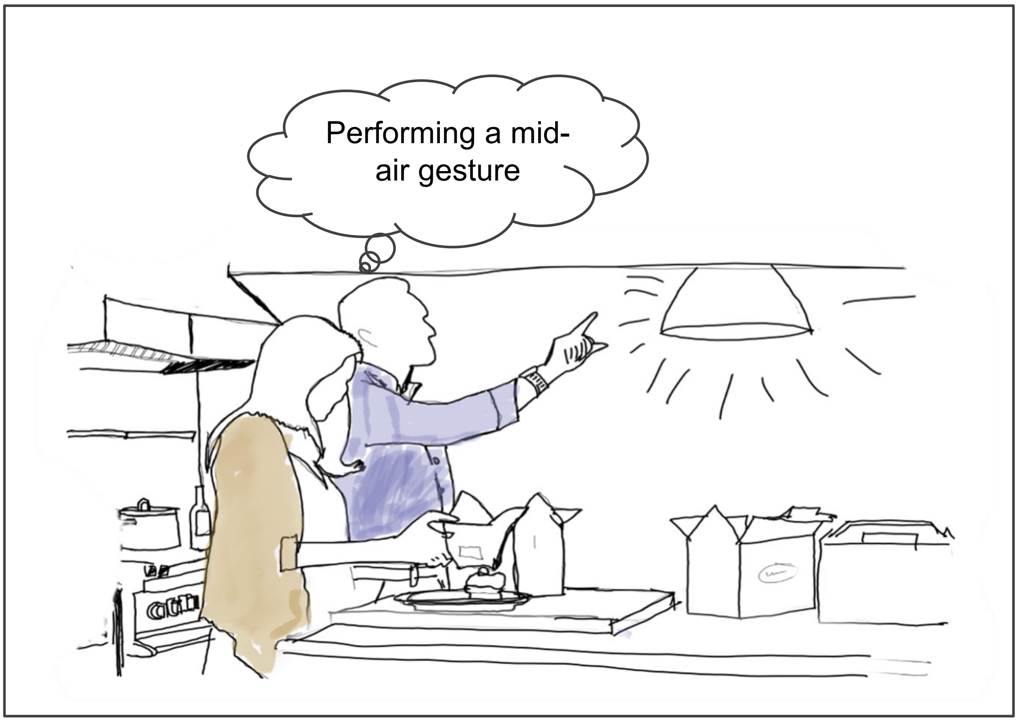



**Figure 4.1:** Illustrative example of a smart home scenario in which users perform mid-air gestures to turn on a kitchen light

In this chapter, we explore how people envision using mid-air gestures to interact with divers devices and in varios scenarios in their smart homes. In the previous chapter, we examined mid-air gesture vocabularies across domains through a literature review, identifying commonalities and proposing a candidate consensus set. In this chapter, we build on those findings by conducting a user study in various smart home scenarios to validate whether consistencies can also be observed when users design gestures in situated smart home scenarios.

Furthermore, existing research has primarily focused on the elicitation of gestures; however, a critical area requiring investigation is the extent to which users want the integration of their devices and associated commands through mid-air gestures within the smart home ecosystem, as well as the motivations that might lead to these preferences. Furthermore, while some studies have considered the realm of multi-device smart home interactions, their singular scenario focus has led to a lack of comprehension of the ways users interact with devices across various real-life scenarios, obstructing the development of user-centric and universally applicable mid-air gesture interfaces.

Furthermore, if we want to incorporate gestures into any device and commands that users want, we should maintain a limited and short set of gestures, as remembering new gestures can be challenging for users. Therefore, it is important to consider the consistency of gestures for similar commands across various devices and scenarios.

To investigate this, we introduced modifications to the basic bottom-up framed guessability approach. Unlike prior studies, which typically exposed participants to system functionality and asked for recommendations for gestures to control that functionality, our approach leverages frames (scenarios) that allow participants to interact with the environment in a manner reflecting their real-life experiences. These scenarios empower us to explore ideas related to motivations and benefits influencing users toward preferring gesture interaction, smart home devices, smart home commands and mid-air gestures associated with each command.

As a result, our methodology encompasses not only gestures but also a wider spectrum of considerations. Our primary focus centers on four distinct user scenarios: cooking, fitness training, TV and entertainment, and home office work. Within these scenarios, in a controlled yet realistic smart home lab setting, we prompted 20 participants to immerse themselves, envisioning the integration of mid-air gestures to control various devices and applications and asking them about their preferences. Furthermore, we employed agreement rates calculation to assess the consensus of gestures across various scenarios and smart home devices. This established measure, originally proposed by Wobbrock et al. [WARM05], allows us to quantify the level of agreement between the various gestures suggested by the participants.

We selected the smart home domain as our focus because it naturally integrates multiple types of interactive devices and diverse user activities. Unlike many single-domain contexts, smart homes bring together scenarios such as cooking, entertainment, fitness training, and home office work, each of which imposes distinct spatial constraints, postures, and attentional demands. These variations provide an opportunity to examine whether and how user preferences remain consistent or change across contexts, while keeping the device ecosystem constant. In addition, the smart home is a domain of ongoing academic exploration and emerging commercial interest for touchless interaction. While voice assistants currently dominate consumer adoption, gesture-based control has been investigated for applications such as lighting, media control, and kitchen appliances.

We first describe in Section 4.1 the study design and analysis approach. We then, in Section 4.2, present results on user motivations, preferences, and consistency across devices and scenarios. Finally, Section 4.3 reflects on the design implications of our results.

Parts of this chapter have been published as a full paper in the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems by Hosseini et al. [HMB24].

## 4.1 Study

To collect and understand the preferences of users regarding mid-air gestures in a smart home environment, we designed a scenario-based elicitation study that would lead users to suggest gestures for various commands and devices within specific scenarios, including cooking, TV and entertainment, home office, and fitness training. The Ethics Committee of our institution gave ethical approval to the study, and informed consent was obtained from all participants.

### 4.1.1 Participants

A total number of 20 people (13 female and 7 male) aged 21-32 (M=25, SD=2.84) participated in our study. The participants did not have any musculoskeletal disorders. Regarding the participants' current living situations, about 75% of the individuals were living in shared flats or with their families, and 25% were living alone. They were recruited through the university's online notice board and our research group environment, and their participation time was reimbursed. All participants were familiar with 2D gesture-based interfaces since all of them own smartphones. Two participants reported having experience with mid-air gestures while playing computer games such as Wii or Kinect, and one participant had conducted research on mid-air gestures in In-Vehicle and Automotive.

### 4.1.2 Apparatus and Setup

Our study took place in a smart home lab at our institution, a fully furnished one-bedroom apartment. It's about 70 $m^2$ and has a standard living area, i.e., living room, bedroom, kitchen, and bathroom. The rooms are equipped with typical household devices and appliances commonly found in a typical German residence, including a television, refrigerator, oven, stove, computer monitor, laptop, mouse, keyboard, speaker, vacuum cleaner, music player, range hood, coffee maker, heater, air conditioner, electronic kettle, washing machine, toaster, blinds, and stationary fitness bike. These devices were placed in various rooms. The key objective of the lab is to design a living space that closely represents the typical characteristics of a German household to enhance participants' immersion in the environment. Therefore, we included additional elements like standard bedroom furniture, office furniture, sofa and entertainment furniture, along with essential kitchen gadgets to ensure the ecological validity of the study. Figure 4.2 visually represents the smart home environment used in our study, showcasing various scenarios. The study was conducted in four different scenarios: cooking, TV and entertainment, home office, and fitness training.

For the *cooking* scenario, participants were positioned standing in the kitchen area in front of the kitchen counter near the oven. Naturally, the kitchen was primarily equipped with typical appliances, including an oven, electric kettle, toaster, stove, range hood, refrigerator, and microwave. Supplementary devices such as smart kitchen gadgets and cooking utensils were also available. This setup allowed participants to interact with kitchen appliances and perform cooking tasks, creating an experience similar to being in their homes. Participants were free to move within the environment.

For the *TV and entertainment* scenario, participants were seated on a comfortable couch in the living room, positioned in front of a television. In this scenario, the TV was the central device. There were also supplementary devices available, including remote controls, smart speakers, and streaming devices.

The *home office* area was thoughtfully designed to replicate a fully functional home office environment. Participants were seated in chairs at a computer workstation. The work desk was equipped with personal computing devices, including laptops, monitors, computer mice, and relevant computer accessories within a home office setting.

In the *fitness training* scenario, the primary device was a stationary fitness bike. Participants were placed on the bike, enabling them to simulate a fitness experience within our Smart Home Lab. Each scenario was conducted in a separate room, except *home office* and *fitness training*, which were co-located in the same room.

Further, in each scenario, we presented participants with illustrations of devices by providing a paper exposing images of the ten most frequently employed smart home appliances, as identified in prior research papers. These included televisions, blinds, ovens, range hoods, appliance control systems, doors, windows, lights, media players, and air conditioners [VK19; KWH+11; SAAG00]. All of these devices were also in our smart home lab. However, participants were not restricted to the listed devices; they were encouraged to introduce their devices if desired. In our study, we refrained from presenting participants with predefined narratives in each scenario. Instead, we encouraged participants to use their imagination and engage with the provided devices and appliances in a manner that felt natural within the given context. This approach allowed participants to act authentically in response to the simulated smart home environments, offering insights into user interactions that were not influenced or biased by predetermined storylines. Participants were asked to envision themselves in each context and interact with the devices as they would in their homes.

### 4.1.3 Procedure

Before the start of the study, participants were welcomed by the experimenters. They were given information about the study, the experimentation setting, and the experiment procedure. Then, an informed consent document was presented to them for reading and signing. They were informed about the tasks that they had to perform. The researchers reported on the types of gestures allowed (performing mid-air gestures with any body part was allowed). We encouraged participants to imagine there were no restrictions on recognizing gestures; any gesture would be recognizable. This approach allowed participants to propose the gestures beyond existing hardware capabilities and offer valuable insights for designing future gesture-based interactions [GVV18]. There was an opportunity for participants to ask questions.

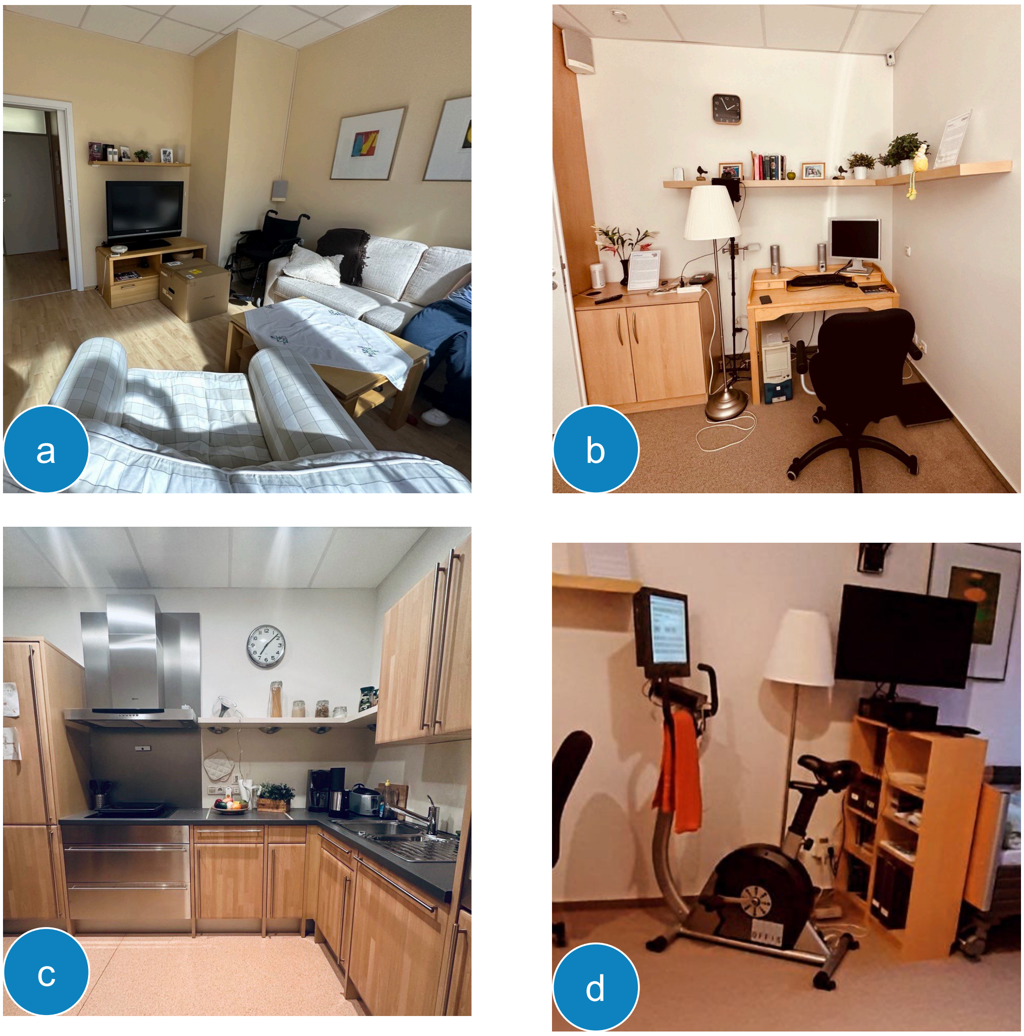



**Figure 4.2:** The visual representation of the smart home environment employed within this study: (a) the living room configuration used for the TV and entertainment scenario, (b) the home office area, replicating a functional workspace for our study's home office scenario, (c) the kitchen area, corresponding to the cooking scenario, (d) the fitness training setup with stationary fitness bikes for the fitness training scenario.

Afterward, participants were given a guided tour of the smart home lab to help them familiarize themselves with the environment and encourage them to explore and engage with the devices described in Section 4.1.2. The experimenter indicated the locations of the devices to the participants. However, to minimize bias, we did not provide detailed information about the functionalities, features, and potential uses of the devices during the tour. Rather, participants were encouraged to interact with the devices, allowing them to independently discover the affordances of the devices. However, the experimenter did respond to participant inquiries. After that, participants were situated in four scenarios in the smart home: cooking, fitness training, TV interaction and entertainment, and home office work. In any scenario:

1. A semi-structured interview was conducted at the beginning of each scenario (e.g., cooking). During this phase, we inquired about their personal experiences, daily habits, and typical activities relevant to the specific scenario.
2. Afterward, participants were tasked to envision themselves in each scenario outlined in Section 4.1.2 (e.g, cooking), where gestures could be used as a means of interacting with household devices and electronic appliances. Here, participants considered the motivations or factors that would prompt them to employ mid-air gestures. They were encouraged to consider specific smart home devices and the corresponding commands for which they would use mid-air gestures. During this phase, we asked them questions about their scenarios in a semi-structured interview, including: what motivated them to use gestural interaction (e.g., dirty hands while cooking); what triggered and motivated their interaction with the devices (e.g., ringing doorbell); and how likely they were to use this scenario in their daily life. We also asked them to show at least one mid-air gesture that they would like to use for the interaction in each scenario for each function that they proposed. To avoid biasing participants towards certain gestures and due to the exploratory nature of the study, no actual gesture input system was present in the study.
3. After finishing all scenarios, participants were interviewed to gain a deeper insight into their preferences regarding mid-air gestural interaction in smart homes, particularly exploring the scenarios, rooms, or commands where they envisioned the most potential for this interaction in their daily lives.
4. Afterward, each participant was asked to answer the questions regarding demographic information and provide feedback regarding their experience with using mid-air gestures. For this, a questionnaire regarding acceptance and cultural aspects (gesture anthropology) of gestural control was proposed by Villanueva and Drögehorn [VD18]. To formulate each question, a 5-point Likert scale was used, where 0 strongly disagreed and 4 strongly agreed. Table 4.1 presents each item and its description. We recorded the entire experiment with a video recording device placed in front of the participants.

We employed a within-subjects design for our study, and the study was conducted in single-user sessions.

### 4.1.4 Factors impacting elicitation

User-generated interaction methods can be affected by legacy bias, in which gestures are proposed by participants based on their prior experiences and interactions. Previous research revealed that the priming technique has the potential to reduce the impact of legacy bias and can produce ecologically valid gestures [MDD+14]. In our study, the priming is achieved with two approaches:

1. Participants were encouraged to reflect on personal experiences and activities that they do in each scenario by asking about personal activities at the beginning of each scenario.
2. By physically placing participants in rooms furnished for specific scenarios and encouraging participants to engage with smart home devices within each scenario.

Further, we conducted this study in a modern yet typical German home, reflective of the environments that many flat and house owners encounter. This implies that appliances and household devices may exhibit inherent affordances, such as knobs on a stove that can be grasped and turned or buttons on a remote control that can be pressed. As cultural context can influence perceptions of affordance [OTC04], we have specifically recruited participants with native German backgrounds to ensure the affordances of devices and the contextual elements within the experiment setting are likely to be well understood. This facilitates a more reliable exploration to see how participants naturally propose gestures in a setting that reflects their everyday lives and to observe behaviors as they naturally occur, without introducing external influences. Considering that affordance is inherently tied to individuals and their cultural context, participants in this study are prone to incorporating pre-existing knowledge related to households and home appliances. The use of this prior knowledge has demonstrated its advantageous impact in fostering more intuitive interactions, thereby enhancing the overall user experience [OTC04].

Additionally, rather than limiting participants to generating a predetermined number of gestures for each command, we allowed individuals to suggest any gesture or body movement that they found fitting, natural, and intuitive. Further, participants were permitted to refine gestures that were proposed earlier.

**Table 4.1:** Likert items on acceptance and culture aspects.

| No. | Statement | Heuristics | Aspect |
|---|---|---|---|
| A1 | Gesture control will improve my overall experience with smart home | Perceived usefulness | acceptance |
| A2 | Gesture control will make interaction with smart home easier | Perceived ease of use | acceptance |
| A3 | I will easily get used to smart home interaction with the help of gestures | attitude towards using the technology | acceptance |
| A4 | Gestures will be a typical way of interacting with technology in the future. | behavior towards intention of use | acceptance |
| C1 | Using gesture is a natural way of interaction with smart home technology | intuitiveness | Cultural |
| C2 | My culture is known to use (hand/body) gestures as part of everyday communication | gesture use | Cultural |
| C3 | My cultural background is known to be very accepting of new technology/innovation. | openness to innovation | Cultural |

### 4.1.5 Data analysis

In this study, the data corpus consists of 700 minutes of video footage, along with questionnaire responses from participants providing their demographics and attitudes toward mid-air gesture usage. The videos were systematically reviewed and coded to extract data. The coding process was inductive and iterative, with the researchers revisiting the videos multiple times to ensure consistency and accuracy in the assigned codes. To enhance the reliability of the coding process, a two-step approach was employed. Initially, two researchers independently evaluated 10% of all data, engaging in a thorough code adjustment session to resolve disagreements through mutual decision. Subsequently, each researcher was assigned all the remaining data, analyzing and coding the information individually. Disagreements between the two researchers were resolved iteratively by revisiting the videos. In this phase, the videos were carefully examined to identify and annotate various elements, including gestures, the use of devices, commands, and verbal explanations provided by participants. All codes were recorded in an Excel file with associated timestamps that describe when and where the coded segments occurred in the videos. The coding process resulted in a total number of 586 gestures, 37 devices, and 60 commands. Each participant, on average, proposed 13 distinct gestures (SD = 4.08). There were conflicts where one gesture was suggested for several commands. We utilized R for data analysis.

Afterward, for further analysis in this chapter, we grouped identical and similar gestures for each command based on the following criteria:

1. Body Part and Handedness: Gestures performed with different body parts were considered similar (e.g. swipe up finger was considered similar to swipe up hand). Additionally, gestures performed with the dominant hand were considered similar to gestures conducted with the non-dominant hand.
2. Size: Gestures executed in different sizes fall into the same category. For example, a "circle" gesture with a large amplitude, performed using the entire arm, is the same as a "circle" gesture with a smaller size performed using the finger.
3. Direction: Gestures with a consistent or identical directionality were considered similar, while those with a different direction were considered different. For example, swiping up and down were considered different.
4. Hand Pose: Gestures that involved different static hand poses were classified into a separate category. For example, a swipe up with the open hand is different from a hand swipe upward with the thumb pointing up.

We ended up with 99 unique gesture categories.

## 4.2 Results

In this section, we present the results of the interactions of smart home users with different devices in different scenarios and the types of mid-air gestures that we observed in and across the different smart home scenarios. First, we present the different motivations that lead users to use this gesture for their gestural interactions and the benefits they see in using gestures in the different scenarios. We then present the preferred devices and the commands that users prefer to control with mid-air gestures. We later analyze the user-generated gestures, analyzing their consistency across different devices and scenarios.

Furthermore, we assess participant attitudes towards mid-air gestural interaction in smart homes.

### 4.2.1 Motivations and benefits

Based on the analysis of users' responses and motivations collected in this study, we identified several categories describing their motivation for gesture-based control. The motivation categories were derived using a bottom-up inductive approach from the data. Figure 4.3 shows the distribution of motivation categories for each scenario. Comfort is one of the most prevalent categories, which is present in substantial proportion in all scenarios, particularly in TV and Entertainment (77%), followed by Home Office (41%). The general motivation for the individuals is to make interaction more accessible and enjoyable from a comfortable position or remotely without having to move. In the Cooking scenario, the majority of motivations mentioned (44%) related to hygienic reasons. Participants highlighted that when cooking their hands were often dirty, greasy, or wet. Mid-air gestures would allow them to avoid touching the surfaces of devices and displays. In the Fitness Training scenario, participants indicated that their primary motivation for using gestures was to reduce interruptions or disruptions to their ongoing tasks or activities, and 49% of participants stated this as their primary motivation. Furthermore, in all scenarios, a portion of the participants expressed that mid-air gestures would provide a more streamlined interaction and could lead to time savings in multitasking.

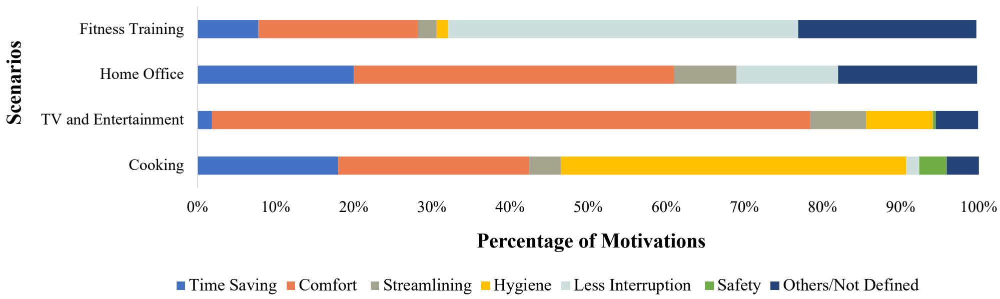



**Figure 4.3:** Distribution of user motivations for using mid-air gesture in different scenarios. The figure shows that the distribution of the motivations varies across the four scenarios. We can observe a strong variation of motivators from safety reasons in the fitness training scenario to hygienic reasons in the cooking scenario.

### 4.2.2 Devices and commands

Our analysis shows that, on average, each participant proposed gestures for approximately 10 devices in the context of the smart home. Unlike previous works that primarily focused on gestures for specific devices, such as TVs [DSSR16; WW12], our study demonstrates that participants want to interact through gestures with a diverse range of devices.

Table 4.2 presents the 14 most frequently mentioned devices, along with the corresponding commands within each scenario. Participants suggested mostly mid-air gestures for various devices, including lights, TVs, blinds, windows, heaters, music players, stoves, ovens, cooker hoods, smartphones, laptops, monitors, air conditioners, and doors, independent of specific scenarios. Furthermore, participants exhibited varying device preferences through mid-air gestures within each scenario. In the Cooking scenario, participants often preferred kitchen appliances, especially the stove, oven, and cooker hood, for gesture interaction, whereas non-appliance devices like lights and smartphones (particularly for music control commands) were also commonly selected. In the TV and Entertainment scenario, the TV emerged as the primary device for mid-air gestures, however, participants also mentioned blinds and light control frequently. In the Home Office scenario, participants mostly mentioned lights and control of blinds. Finally, in the Fitness Training scenario, the laptop was frequently selected (particularly for music control commands); afterward, controlling lights and air conditioning were mentioned more than other devices.

Additionally, our analysis revealed that participants preferred designing gestures for common and fundamental tasks on these devices. For instance, they mostly suggested gestures for functions such as turning devices on or off and adjusting basic settings like increasing or decreasing different aspects such as volume, temperature, and lighting. Several participants shared their viewpoints on managing complex tasks, expressing a preference for alternative interaction modalities like remote controls over mid-air gestures in these cases. For instance, a participant noted, *commands like selecting specific items from a TV using gestures can be quite challenging for me. Consequently, I prefer using remote controller devices instead of relying on gestures for such activities. (P19)*

**Table 4.2:** Device types, usage scenarios, and corresponding commands

| Devices | Scenarios | Commands |
|---|---|---|
| Light (80) | Cooking (15) | Turn on(10), turn off(3), darker(1), Change Color (1) |
| | TV and Entertainment (32) | Turn on(13), turn off(9), darker(5), brighter(3), Change Color (2) |
| | Home Office (24) | Turn on(12), turn off(5), darker(3), brighter(3), Change Color (1) |
| | Fitness Training (9) | Turn on(5), turn off(1), darker(2), brighter(1) |
| TV (76) | TV and Entertainment (74) | Next(7), Previous(7), Regulate Volume, Select(3), Turn Off(10), Turn On(12), Volume Down(16), Volume Up(15), Standby Mode (1), pose(1), play(1), Open Menu(1) |
| | Fitness Training (2) | Fast Forward(1), Rewind(1) |
| Blinds (61) | Cooking (7) | Move Up(4), Move Down(3) |
| | TV and Entertainment (27) | Move Up(12), Move Down(15) |
| | Home Office (19) | Move Up(8), Move Down(11) |
| | Fitness Training (8) | Move Up(4), Move Down(4) |
| Window (47) | Cooking (8) | Open(5), Close(2), Tilt Mode(1) |

*Continued on next page*

**Table 4.2 – continued from previous page**

| Devices | Scenarios | Commands |
|---|---|---|
| | TV and Entertainment (12) | Open(4), Close(4), Tilt Mode(4) |
| | Home Office (14) | Open(6), Close(6), Tilt Mode(2) |
| | Fitness Training (13) | Open(6), Close(4), Tilt Mode(3) |
| Heater (41) | Cooking (2) | Temperature Up(1), Temperature Down(1) |
| | TV and Entertainment (18) | Temperature Up(8), Temperature Down(8), turn on(1), turn off(1) |
| | Home Office (12) | Temperature Up(5), Temperature Down(5), turn on(2) |
| | Fitness Training (9) | Temperature Up(4), Temperature Down(3), turn on(1), turn off(1) |
| Music Player (35) | Cooking (5) | Volume Up(2), Volume Down(1), Turn on(1), Turn off(1) |
| | TV and Entertainment (13) | Volume Up(3), Volume Down(3), Turn on(4), Turn off(1), Next(1), Play(1) |
| | Home Office (3) | turn on(2), turn off(1) |
| | Fitness Training (14) | Volume Up(4), Volume Down(3), Turn on(2), Turn off(1), pause(1), Next(2) |
| Stove (28) | Cooking (28) | Turn Off(8), Turn On(7), Temperature Down(5), Temperature Up(6), Select(1), Set Timer(1) |
| Oven (22) | Cooking (22) | Turn Off(4), Turn On(5), Temperature Down(3), Temperature Up(3), Last Setting Recall (1), Close(1), Convection (1), Open(2), Pull out tray (1), choose programme(1) |
| Cooker Hood (22) | Cooking (22) | Turn On(13), Turn Off(4), Regulate Light(1), Speed Up(1), Temperature Down(1), Temperature Up(1), Turn on light(1) |
| Smartphone (19) | Cooking (11) | Pause(3), Play(2), Next(2), Screen On(1), Scroll(1), Volume Down(1), Volume Up(1) |
| | TV and Entertainment (1) | Tell the Time(1) |
| | Home Office (2) | Play(1), pause(1) |
| | Fitness Training (5) | Next(1), Pause(1), Play(1), X repetitions more of the exercise(1), X repetitions less of the exercise(1) |
| Laptop (16) | TV and Entertainment (3) | Next(1), Previous(1) |

*Continued on next page*

**Table 4.2 – continued from previous page**

| Devices | Scenarios | Commands |
|---|---|---|
| | Home Office (2) | Move between pages(1), Save(1) |
| | Fitness Training (11) | Pause(3), Next(2), Play(2), Mute(1), Previous(1), Volume Down(1), Volume Up(1) |
| | Cooking (1) | Turn On(1) |
| | Home Office (14) | Move between pages(2), Scroll(1), Select(1), Turn Off(1), Turn On(5), Zoom(4) |
| | Fitness Training (3) | Switch through program(1), Turn Off(1), Turn On(1) |
| Air Conditioner (1 | TV and Entertainment (3) | Temperature Down(1), Temperature Up(1), Turn On(1) |
| | Home Office (3) | Temperature Down(1), Temperature Up(1), Turn On(1) |
| | Fitness Training (8) | Temperature Down(3), Temperature Up(3), Turn Off(1), Turn On(1) |
| Door (10) | TV and Entertainment (2) | Close(2) |
| | Home Office (5) | Close(3), Open(2) |
| | Fitness Training (3) | Close(2), Open(1) |

### 4.2.3 Gesture type

While participants were free to use any body part, they most commonly used their upper limbs to perform the gestures (91%), with hand and fingers (46%) followed by the arm (28%) and forearm (17%). About 8% of participants used a combination of hand and arm or hands and forearms. Only 1% of the suggested gestures involved the lower limbs (foot and leg). Regarding manual gestures, uni-manual gestures (82%) were more common than bi-manual gestures (18%). Uni-manual gestures were performed with either the right hand (66%) or the left hand (34%). The intensity of the gestures depended on both the participant's hand dominance and the device's position relative to the participant. Among the bi-manual gestures, the majority (91%) were symmetrical.

Across these body parts, participants exhibited preferences for various gesture types, such as swiping in different directions, including upward (10%), downward (11%), left/right (12%), and forward and backward, with (4%) for pushing forward and (5%) for pulling backward. Additionally, participants frequently used clapping hands (5%) and rotating the hand in a manner similar to turning a knob (7%).

We refrained from specifying the direction for 'pointing' gestures and also 'swipe left' and 'swipe right.' Instead, we grouped 'swipe left' and 'swipe right' together as 'swipe left/right' and when it was pointing to the device, without mentioning its direction, due to their frequent correlation with the orientation of the device and the position of the participant relative to the device.

### 4.2.4 Agreement and user-defined gesture set

To determine the consensus level of the participants' suggestions, we calculated the agreement rate for each suggestion based on the agreement rate introduced by Vatavu and Wobbrock [VW15]:

$$AR(c) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} (\frac{|P_i|}{|P|})^2 - \frac{1}{|P|-1} \tag{4.1}$$

In Equation (1), P represents the set of whole gestures proposed for command c, and $P_i$ is the $i^{th}$ subset of gestures that are unique in P. We calculated the agreement rate for all commands and classified agreement as very high (AR>0.5), high (0.3<OR<0.5), medium (0.1<AR<0.3), and low (0.1<AR) [VW15]. Table 4.3 shows the results. In this table, the first column indicates commands along with the number of times they were mentioned by all participants across all scenarios. The second column shows the agreement rate, and the third column lists the two most frequent gestures associated with each command. The last column presents the number of unique gestures associated with each command. Whenever two gestures for a command had the same number of repetitions, we selected the gesture presented by more participants than gestures proposed multiple times by the same participant.

Based on our analysis, agreement rates varied from 0 to 1, and the mean agreement across all commands and devices was 0.3 (SD= 0.33). A total of 8 commands were rated as very high in terms of agreement, 4 as high, 9 as medium, and 5 as low. High or very high agreement was found for the following commands: *speed up*, *move between pages*, *previous*, *scroll*, *zoom*, *next*, *move up*, *move down*, *tilt mode*, *close*, *open*, and *volume up*. On the other hand, the commands that were rated as low or medium agreement were: *brighter*, *temperature down*, *volume down*, *temperature up*, *select (UI element)*, *change color*, *darker*, *standby mode*, *turn on*, *turn off*, *pause/stop*, *play*, *set timer*, *switch to a different setting*. While high consistency was observed for some commands, such as *speed up*, and consistency was not achieved for some commands, like *set timer*, it is important to note that the number of proposed gestures for these commands was limited (2 instances for each command), which may impact the generalizability of these results.

After identifying identical gestures for each command and calculating the agreement rate, we define a user-defined gesture set. For this, we selected 2 of the largest groups of identical gestures for each command. Most of the highly recommended gestures (85%) are basic anatomical gestures, such as a single tap, a swiping motion, or a point. In contrast, some of the other commonly suggested gestures are constructed by combining two distinct gestures, such as pointing followed by the motion of turning the knob. Additionally, dynamic gestures, such as clapping hands or swiping, constituted the majority of proposed gestures (91.4%), while static gestures, such as pointing or displaying numerical values using fingers, constituted only 8.6% of these gestures.

**Table 4.3:** Agreement rate for each command on different devices. The fields show the agreement rates for each device/appliance and gesture. The colors (when viewed in color) range from red with a low agreement rate to green with a high agreement rate. For example, the heatmap shows a high agreement rate for opening and closing the fridge but low agreement rates for turning off the TV. The figure also offers an overall agreement rate across devices.

| **Command** | **AR** | **Top two Gestures** | ***Distinct Gestures*** |
|---|---|---|---|
| Turn On (125) | 0.10 | Clap Hands (27), Finger Snapping (23) | 28 |
| Turn Off (61) | 0.08 | Clap Hands (13), Finger Snapping (10) | 23 |
| Open (34) | 0.33 | Pull Hand Back (19), Swipe Left/Right (6) | 10 |
| Move Down (34) | 0.58 | Swipe Down (26), Grab Hand and Move it Down (2) | 6 |
| Temperature Up (33) | 0.22 | Turning Knob (12), Swipe Up (9) | 8 |
| Temperature Down (31) | 0.28 | Turning Knob (11), Swipe Up (9) | 7 |
| Volume Up (31) | 0.31 | Swipe Up (17), Turning Knob (4) | 12 |
| Close (30) | 0.40 | Push Hand Forward (18), Swipe Left/Right (7) | 8 |
| Volume Down (30) | 0.28 | Swipe Down (15), Turning Knob (4) | 12 |
| Move Up (27) | 0.60 | Swipe Up (21), Point to the Device (2) | 6 |
| Next (17) | 0.67 | Swipe Left/Right (14), Turning Knob (1) | 4 |
| Darker (12) | 0.11 | Move the Fingers Together (3), Swipe Down (3) | 7 |
| Pause/Stop (11) | 0.05 | Move Hand Away from Ear (2), Tap (2) | 8 |
| Tilt Mode (10) | 0.47 | Pull Hand Back (7), Rotation (1) | 4 |
| Previous (9) | 1.00 | Swipe Left/Right (9) | 1 |
| Play (9) | 0.03 | Move Hand Away from Ear (2), Tap (1) | 8 |
| Brighter (8) | 0.21 | Move the Fingers Away from the Thumb (3), Swipe Up (3) | 4 |
| Move Between Pages (5) | 1.00 | Swipe Left/Right (5) | 1 |
| Standby Mode (5) | 0.10 | Swiping Diagonally (2), Sleeping Position (1) | 4 |
| Select (UI element) (4) | 0.17 | Point to the Device (2), Point and Double Tap (1) | 4 |
| Change Color (4) | 0.17 | Finger Snapping Twice (2), Point then Turning Knob (1) | 3 |
| Zoom (3) | 1.00 | Fingers Closing/Opening (3) | 1 |
| Speed Up (2) | 1.00 | Swipe Up (2) | 1 |
| Scroll (2) | 1.00 | Swipe (2) | 1 |
| Set Timer (2) | 0.00 | Point then Turning Knob (1), Tap on the Wrist then Show Number with Fingers (1) | 2 |

*Continued on next page*

**Table 4.3 – continued from previous page**

| **Command** | **AR** | **Top two Gestures** | ***Distinct Gestures*** |
|---|---|---|---|
| Switch to a Different Setting (2) | 0.00 | Point to the Device Twice (1), Show Number with Finger (1) | 2 |

### 4.2.5 Gesture consistency across devices

We calculated the agreement rate for each command across different devices separately and in a cross-device situation, meaning independent of the devices. Figure 4.4 displays a heatmap plot of the findings. Here, red shows medium and low agreement, and green shows very high and high agreement. The heatmap in Figure 4.4 does not include commands if there is only one gesture per command in the cross-device situation (i.e., too few gestures to calculate AR). Commands without gestures on a device are denoted by “NA”, and commands with only one gesture on a device are denoted by “*”.

Most commands with high or very high agreement in cross-device situations also have high or very high agreement on individual devices. The exception to this is *next* on smartphone (AR=0), *open* on oven (AR=0), and *volume up* on TV (AR=0.09). Furthermore, the majority of the commands with low or medium agreement in cross-device situations also have low or medium agreement in individual devices, with the exception of *turn off*, which has high agreement in cooker hood (AR=0.5), *turn on*, which has high agreement in computer case and water tap (respective AR=0.33 and AR=0.4 respectively), *volume down*, which has very high agreement in music player (AR=0.52), *temperature down* in heater and stove (AR=0.38 and AR=0.4), and *temperature up*, which is high in heater (AR=0.45), and *select (UI element)* in TV (AR=0.33).

In addition, to determine whether the most commonly recommended gestures for each device with high AR remain consistent compared to the general user-defined gesture set outlined in table 4.3, we identified the most commonly recommended gesture for each command on each device with high AR. The outcomes of this analysis are presented in Table 4.4. It can be observed that there is a tendency among the participants to employ consistent types of gestures across various devices for similar commands. For instance, the *open* command commonly involves swiping to the left or right or pulling the hand back, depending on whether it is related to a window, door, cupboard, or refrigerator.

Similarly, the *close* command often involves pushing the hand forward or swiping to the left or right. The commands *next* and *previous* typically entail swiping to the left or right on devices such as TVs, music players, and laptops. Furthermore, binary commands that involve two states, such as on/off or open/close, frequently exhibit either the same gesture or opposite gestures. For example, the *next* and *previous* commands across TVs, music players, and laptops are consistently associated with swiping left or right. Similarly, the gestures for *temperature up* and *temperature down* in different devices involve either rotation or swiping up and down. However, for certain commands, the choice of the most frequent gesture was highly dependent on the specific devices involved. Examples of such commands include *turn on*and *turn off*.

Furthermore, during our study, we observed that most of the users were interested in the potential of a universal gesture set. It was observed that many participants showed a tendency towards the idea of reusing the same gesture across different devices for the same command by executing the gesture towards the device direction.

**Table 4.4:** The most commonly proposed gestures for commands on devices with high AR.

| Command | Device and the most common gesture |
|---|---|
| Open | **Window:** Pull Hand Back (12 of 21); **Cupboard:** Pull Hand Back (3 of 4); **Door:** Swipe Left/Right (3 of 3); **Fridge:** Pull Hand Back (2 of 2) |
| Close | **Window:** Push Hand Forward (12 of 16); **Door:** Swipe Left/Right (6 of 7); **Cupboard:** Push Hand Forward (2 of 2); **Fridge:** (Push Hand Forward (2 of 2) |
| Temperature Down | **Heater:** turning knob(10 of 17); **Stove:** Swipe Down (3 of 5) |
| Temperature Up | **Heater:** turning knob (11 of 18); |
| Turn On | **Water Tap:** Wave (Left and Right) (3 of 5); **Computer Case:** Hold the hand near to the device (2 of 3) |
| Turn Off | **Cooker Hood:** Swipe Left/Right (3 of 4) |
| Volume Up | **Music Player:** Swipe Up (7 of 9); |
| Volume down | **Music Player:** Swipe down (5 of 7); |
| Move Between Pages | **Monitor:** Swipe Left/Right (2 of 2) |
| Select (UI element) | **TV:** Point (2 of 3) |
| Next | **Music Player:** Swipe Left/Right (2 of 3); **Laptop:** Swipe Left/Right (3 of 3); **TV:** Swipe Left/Right (7 of 7) |
| Previouse | **Laptop:** Swipe Left/Right (2 of 2); **TV:** Swipe Left/Right (7 of) |
| Zoom | **Monitor:** Fingers Closing/Opening (4 of 4) |
| Move Up | **Blinds:** Swipe Up (21 of 27) |
| Move Down | **Blinds:** wipe Down (26 of 34) |
| Tilt Mode | **Window:** Pull Hand Back (7 of 10) |

### 4.2.6 Gesture consistency across scenarios

To determine how the agreement rate varies for individual commands in different scenarios, we calculated the agreement rate for each command within each scenario. Figure 4.5 presents the agreement rates across various scenarios for each command. In cross-scenario situations, i.e., independent of the scenarios, the majority of commands that exhibit either high or very high AR also exhibit high AR within individual scenarios. However, there were some exceptions. For example, *tilt mode* in TV and entertainment has medium (AR=0.17), *close* in TV and entertainment, home office, and fitness training has medium (AR=0.27, AR=0.29, AR=0.27 respectively), *open* in TV and entertainment and fitness training has medium (AR=0.17 and AR=0.14 respectively), and *volume up* in TV and entertainment has medium (AR=0.14).

Furthermore, most commands with lower agreement rates in cross-scenario situations display low agreement rates within different scenarios as well. Nevertheless, there were

**Devices**

| Command | Cross-Device | Air Conditioner | Light | TV | Oven | Music Player | Smartphone | Monitor | Heater | Stove | Laptop | Fridge | Cooker Hood | Window | Blinds | Headphone | Door | Computer Case | PC | Printer | Cupboard | Fitness Equipment | Coffee Machine | Tablet | Toaster | Vacuum cleaner | Water Tap | Curtain | Kettle | Book Reader | Microwave |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speed Up | 1 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Move Between Pages | 1 | NA | NA | NA | NA | NA | NA | 1 | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | * | NA |
| Previous | 1 | NA | NA | 1 | NA | NA | NA | NA | NA | NA | 1 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Scroll | 1 | NA | NA | NA | NA | NA | * | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Zoom | 1 | NA | NA | NA | NA | NA | NA | 1 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | * | NA |
| Next | 0.67 | NA | NA | 1 | NA | 0.33 | 0 | NA | NA | NA | 1 | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA |
| Move Up | 0.6 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.6 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Move Down | 0.58 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.58 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Tilt Mode | 0.47 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 0.47 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Close | 0.4 | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | 1 | NA | 0.55 | NA | NA | 0.71 | NA | NA | NA | 1 | NA | NA | NA | NA | NA | NA | * | NA | NA | * |
| Open | 0.33 | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | 1 | NA | 0.32 | NA | NA | 1 | NA | NA | NA | 0.5 | NA | NA | NA | NA | NA | NA | NA | NA | NA | * |
| Volume Up | 0.31 | NA | NA | 0.09 | NA | 0.61 | * | NA | NA | NA | * | NA | NA | NA | NA | * | NA | NA | * | NA | NA | * | NA | * | NA | NA | NA | NA | NA | NA | NA |
| Brighter | 0.21 | NA | 0.21 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Temperature Down | 0.28 | 0 | NA | NA | 0 | NA | NA | NA | 0.38 | 0.4 | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Volume Down | 0.23 | NA | NA | 0.09 | NA | 0.52 | * | NA | NA | NA | * | NA | NA | NA | NA | * | NA | NA | * | NA | NA | * | NA | * | NA | NA | NA | NA | NA | NA | NA |
| Temperature Up | 0.22 | 0 | NA | NA | 0 | NA | NA | NA | 0.45 | 0.27 | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Select (UI element) | 0.17 | NA | NA | 0.33 | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Change Color | 0.17 | NA | 0.17 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Darker | 0.11 | NA | 0.11 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Standby Mode | 0.1 | NA | NA | * | NA | NA | NA | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Turn On | 0.1 | 0 | 0.29 | 0.12 | 0.1 | 0.11 | NA | 0.05 | 0 | 0 | NA | NA | 0.17 | NA | NA | NA | NA | 0.33 | NA | 0 | NA | NA | 0 | NA | * | * | 0.4 | NA | 0 | NA | NA |
| Turn Off | 0.08 | * | 0.27 | 0.04 | 0 | 0 | NA | 0 | 0 | 0.04 | NA | NA | 0.5 | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | * | 0 | NA | NA | NA | NA |
| Pause/Stop | 0.05 | NA | NA | * | NA | * | 0.1 | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Play | 0.03 | NA | NA | * | NA | * | 0.17 | NA | NA | NA | 0 | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Set Timer | 0 | NA | NA | NA | NA | NA | NA | NA | NA | * | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Switch to a different setting | 0 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | * | * | NA | NA | NA | NA | NA | NA | NA | NA |

0 1

**Figure 4.4:** Agreement rate for each command on different devices. The fields show the agreement rates for each device/appliance and gesture. The colors (when viewed in color) range from red with a low agreement rate to green with a high agreement rate. For example, the heatmap shows a high agreement rate for opening and closing the fridge, but low agreement rates for turning off the TV. The figure also offers an overall agreement rate across devices.

some exceptions. For example, *temperature down* has a low agreement rate in the cross-scenario situation but achieves a high agreement rate in the TV and entertainment scenario (AR=0.33). *Temperature up* has high AR in the TV and entertainment scenario (AR=0.33) and in the fitness training scenario (AR=0.33). *Volume down* has a very high agreement rate in the home office scenario (AR=1) and the fitness training scenario (AR=0.67). *Brighter* and *darker* have high agreement rates in the home office scenario (AR=0.33 and AR=0.33 respectively). *Turn off* has a high agreement rate in fitness training (AR=0.3).

Additionally, to determine whether the most commonly recommended gesture remains consistent across scenarios with a high degree of agreement, we identified the most frequently suggested gesture for each command within every scenario. Our focus was on commands and scenarios that exhibit high or very high agreement. Table 4.5 shows the results. We can see that many users tend to use identical gestures for similar commands in different scenarios. The gesture that occurs most frequently within individual scenarios is also the first or, at least, the second most common gesture in cross-scenario situations for every command. For instance, in both scenarios of cooking and home office, the *open*

**Table 4.5:** The most commonly proposed gestures for commands on scenarios with high AR.

| **Command** | **Scenario and most frequent gesture** |
|---|---|
| Open | **Cooking:** Pull Hand Back (10 of 15), **Home Office:** Pull Hand Back (4 of 8) |
| Close | **Cooking:** Push Hand Forward (7 of 8) |
| Volume Up | **Cooking**: Swipe Up (2 of 3), **Home Office:** Swipe Up (2 of 2), **Fitness Training:** Swipe Up (6 of 7) |
| Brighter | **Home Office:** Swipe Up (2 of 3) |
| Temperature Down | **TV and entertainment:** Turning knob (5 of 9) |
| Volume Down | **Home Office:** Swipe Down (2 of 2), **Fitness Training:** Swipe Down (5 of 6) |
| Temperature Up | **TV and entertainment:** Turning knob (5 of 9), **Fitness Training:** Turning knob (4 of 7) |
| Select (UI element) | **TV and entertainment:** Point (2 of 3) |
| Darker | **Home Office:** Swipe Down (2 of 3) |
| Turn Off | **Fitness Training:** Finger Snapping (3 of 5) |
| Move Between Pages | **Home Office:** Swipe Left/Right (3 of 3) |
| Previous | **TV and entertainment:** Swipe Left/Right (8 of ) |
| zoom | **Home Office:** Fingers Closing/Opening (3 of 3) |
| Next | **TV and entertainment:** Swipe Left/Right (8 of 9), **Fitness Training:** Swipe Left/Right (4 of 6) |
| Move Up | **Cooking:** Swipe Up (2 of 3), **TV and entertainment:** Swipe Up (8 of 12), Home Office: Swipe Up (8 of 8), **Fitness Training:** Swipe Up (3 of 4) |
| Tilt Mode | **Home Office:** Pull Hand Back (2 of 2), **Fitness Training:** Pull Hand Back (2 of 3) |

command is frequently associated with the pull hand back gesture, which is also similar to the most frequent gesture identified for this command in Table 4.3.

Table 4.6 presents a set of gestures observed across multiple scenarios. Among all the identified gestures, swipe in different directions (e.g., left, right, up, and down), rotate the hand, clap hands, horizontally wave, point to the device, tap, pull hand back, push hand forward, and finger snapping are gestures that occurred across all scenarios. Some gestures appeared in only two or three scenarios. As an example, 'show palm' appeared in the entertainment and fitness training scenarios, as well as 'move hand away from ear' that appeared in the cooking and home office scenarios.

### 4.2.7 Participant attitudes towards mid-air gestural interaction

During the post-study interviews, participants expressed their most preferred scenarios for using gestures in their daily life. A majority of participants indicated a strong inclination towards utilizing gestural interaction primarily in the kitchen (48%), followed by (35%)

**Command**

| Scenario | Speed Up | Move Between Pages | Previous | Scroll | Zoom | Next | Move Up | Move Down | Tilt Mode | Close | Open | Volume Up |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cross-Scenario | 1 | 1 | 1 | 1 | 1 | 0.67 | 0.6 | 0.58 | 0.47 | 0.4 | 0.33 | 0.31 |
| Cooking | * | * | | * | NA | * | 0.33 | 1 | * | 0.75 | 0.43 | 0.33 |
| TV and entertainment | NA | * | 1 | NA | * | 0.78 | 0.42 | 0.43 | 0.17 | 0.27 | 0.17 | 0.14 |
| Home Office | NA | 1 | NA | * | 1 | * | 1 | 0.65 | 1 | 0.29 | 0.32 | 1 |
| Fitness Training | * | NA | * | NA | NA | 0.4 | 0.5 | 0.5 | 0.33 | 0.27 | 0.14 | 0.71 |

| Scenario | Brighter | Temperature Down | Volume Down | Temperature Up | Select (UI element) | Change Color | Darker | Standby Mode | Turn On | Turn Off | Pause/Stop | Play | Set Timer | Switch to a different setting |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cross-Scenario | 0.21 | 0.28 | 0.23 | 0.22 | 0.17 | 0.17 | 0.11 | 0.1 | 0.1 | 0.08 | 0.05 | 0.03 | 0 | 0 |
| Cooking | NA | 0.27 | 0 | 0.22 | NA | * | * | NA | 0.08 | 0.07 | 0 | 0 | 0 | * |
| TV and entertainment | 0 | 0.33 | 0.13 | 0.33 | 0.33 | 0 | 0 | 0 | 0.16 | 0.12 | * | 0 | NA | NA |
| Home Office | 0.33 | 0.27 | 1 | 0.27 | * | * | 0.33 | 0 | 0.04 | 0.11 | * | * | NA | NA |
| Fitness Training | * | 0.27 | 0.67 | 0.33 | NA | NA | 0 | * | 0.2 | 0.3 | 0.07 | 0 | NA | * |

0 — 1

**Figure 4.5:** Agreement rate for each command on different scenarios. This indicates whether users would use the same commands in different scenarios.

during TV and entertainment. Conversely, home office and fitness training scenarios exhibited considerably lower demand, with 13% and 4% of participants, respectively. Participants provided explicit reasons for their preference for the kitchen scenario, emphasizing its practicality in controlling frequently used appliances and equipment, particularly in multitasking situations. Additionally, they noted the convenience of touch-free interaction, which allowed for the maintenance of hygiene while controlling devices. On the other hand, for the TV and entertainment scenario, participants highlighted the importance of convenience and the ability to manage devices from a distance without having to relocate from their current position.

Questions regarding the number of distinct gestures wanted for use throughout the smart home revealed a wide range of preference. A significant proportion of participants (37%), leaned towards a preference for 5 to 9 distinguishable gestures. In contrast, 32% favored fewer, setting a cap of 5 gestures as their preferred maximum limit. Essentially, their motivation revolved around ease of memorizing and avoiding confusion and complexity. Notably, 26.32% of participants expressed a desire for a more extensive and enriched gestural interaction experience, specifying a range of 10 to 15 distinguishable gestures.

Participants' responses regarding their preferences on the integration of smart home technology into specific rooms of their homes can be categorized into four distinct groups. Approximately 87% of participants indicated that privacy concerns would prevent them from using such technology in their bathroom and bedroom. On the other hand, 25% of participants were open to having gestural interaction technology in every room without any particular concerns regarding how it would be implemented. Six percent of participants prioritized the living room as a place of relaxation and favored minimal technological intervention. Finally, 6% of participants mentioned that their preferences depended on the type of technology used. For example, they were open to certain technologies in specific rooms but had concerns about camera-based tracking systems.

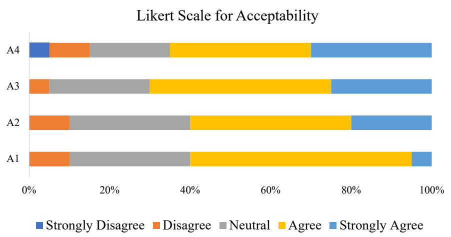
### 4.2.8 Acceptance and cultural aspects of using gesture control

Figure 4.6 presents the results of Likert scale assessments regarding the acceptance and cultural aspects of performing mid-air gestures. In the acceptance category, an average score of 2.7 out of 4 was obtained. Notably, statement A3, *"I will easily get accustomed*

**Table 4.6:** Gestures recurring in various scenarios with repetition of each gesture in each scenario

| | Cooking | TV and Entertainment | Home Office | Fitness Training |
|---|---|---|---|---|
| Swipe Left/Right | 23 | 26 | 12 | 11 |
| Swipe Down | 15 | 20 | 15 | 12 |
| Rotation | 5 | 19 | 8 | 10 |
| Clap Hands | 13 | 19 | 7 | 6 |
| Swipe Up | 13 | 18 | 14 | 12 |
| Finger Snapping | 9 | 16 | 5 | 8 |
| Wave (Left and Right) | 7 | 5 | 4 | 2 |
| Point to the Device | 3 | 5 | 2 | 1 |
| Pull Hand Back | 12 | 4 | 6 | 7 |
| Push Hand Forward | 13 | 7 | 7 | 6 |
| Tap | 5 | 4 | 2 | 3 |
| Move Both Arms Towards Each Other | 1 | 4 | 2 | 0 |
| Clap Hands Twice | 2 | 8 | 1 | 0 |
| Move Both Arms Away from Each Other | 0 | 3 | 1 | 1 |
| Hold the hand near to the device | 2 | 2 | 3 | 0 |
| Point to the Device Twice | 1 | 6 | 0 | 0 |
| Point then Rotation | 1 | 3 | 2 | 0 |
| Grab hand and Pull it Back | 1 | 2 | 1 | 0 |
| Move the Fingers away | 0 | 1 | 1 | 1 |
| Move the Fingers Together | 0 | 1 | 1 | 1 |
| Finger Snapping Twice | 1 | 1 | 1 | 0 |
| Grab Hand and Move it Down | 1 | 1 | 1 | 0 |
| Swiping Diagonally | 0 | 1 | 1 | 0 |
| Show Palm | 0 | 1 | 0 | 3 |
| Show Number with Fingers | 1 | 0 | 4 | 1 |
| Move Hand Away from Ear | 2 | 0 | 2 | 0 |
| grab hand and Push it Forward | 0 | 0 | 1 | 1 |
| Point then Finger Snapping | 0 | 0 | 1 | 1 |
| Stomping Foot on the Ground | 2 | 0 | 0 | 2 |
| Cross the Arms | 1 | 0 | 0 | 1 |
| Fanning Yourself | 1 | 0 | 0 | 1 |

*to smart home interaction with the assistance of gestures,"* garnered the highest mean score within this category (mean = 2.9, SD = 0.85). An average score of 2.5 out of 4 was observed in the cultural category. Statement C2, *"My culture is known to incorporate (hand/body) gestures as a part of everyday communication,"* received the highest mean score within this category (mean = 3.15, SD = 1.13).



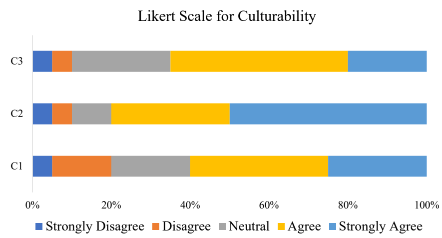




**Figure 4.6:** This diagram illustrates the percentage of participants who provided responses ranging from strong disagreement (0 on the Likert scale) to strong agreement (4 on the Likert scale) to the questionnaire related to acceptance and culture in the context of using mid-air gestures.

## 4.3 Discussion

This chapter aimed to identify user preferences for mid-air gestures, devices, and commands for interaction in different smart home scenarios. It also aimed to investigate the degree of consistency in participants' preferences and recommendations when transitioning to alternative scenarios in the smart home context. To consider this, we proposed modifications to basic elicitation approaches by adapting the bottom-up frame-based design process to encompass a broader exploration of user engagement with smart home technologies. We designed scenarios that allowed participants to interact with the environment in a way that mirrored their real-life experiences. We explored the motivations and benefits influencing users' preference for gesture interaction and considered relevant aspects of gestural interaction.

In contrast to previous elicitation studies where participants were typically exposed to system functionality and asked to recommend gestures for control, our approach utilized framed scenarios to enable participants to elicit ideas concerning smart home devices and functionalities, thus incorporating not only gestures but a broader spectrum of smart home interaction. In the following section, we provide the key takeaways derived from our study.

### 4.3.1 Consistency and User Preferences

Participants exhibited high consistency in employing specific gestures for navigation commands, including *next*, *previous*, *move between pages*, *zoom*, or *scroll* using prevalent gestures such as swiping left or right, moving fingers apart or together in different scenarios. This consistency aligns with the widespread use of these navigation actions in commercial gestural interfaces. Therefore, the observed consistency in these commands suggests legacy bias influenced by participants' prior experiences with similar digital interfaces.

Despite studies arguing for the removal of legacy bias [MDD+14], as it can directly affect the proposed gestures and may limit the potential for producing gestures that fully utilize emerging technologies and sensors, other studies observed that retaining these familiar or legacy gestures had a positive impact on how users perceived and interacted with new technology. Other studies also suggested that legacy bias is useful for researchers to discover gestures that are consistent and, therefore, transferable across various types of devices [KB15; VW22]. By ignoring legacy biases, we force participants to learn new concepts that may not be intuitive to them. For example, previous research has demonstrated that participants perceive swipe gestures as highly usable for navigation commands, considering factors such as physical and mental effort demands for their intended use [KD16]. Thus, rather than dismissing legacy gestures outright, it is essential to take into account other factors that affect the user experience.

We also observed higher consistency among participants for proposed gestures for certain commands that relate to an intended physical change of a physical object, such as 'move blinds up or down.' In these cases, most participants favored gestures that closely aligned with the spatial and directional aspects of the corresponding physical commands. Therefore, participants can find a simple mapping for these kinds of commands relatively easily. For instance, the majority of participants suggested swiping their hand upward to signify moving the blinds up. This consensus was found to be independent of the specific scenario. This is in line with the principles of spatial mapping in Human-Computer Interaction (HCI)[PO15]. We observed the effective application of spatial mapping in mid-air gestural interactions within the smart home environment, and it demonstrated high consistency.

Furthermore, we observed that among the different proposed gestures, some are widget-specific (metaphoric gestures), such as turning a knob for many commands. For instance, when it comes to commands related to temperature control (up or down), this gesture was identified as the most frequently used. The command of turning a knob, when directly related to the device (where a widget or knob is visibly present on the device), was expected to exhibit similar levels of consistency for the same command across different devices. However, as observed in our investigation of heaters and ovens, both equipped with a knob in our home devices, the consistency for changing temperature varied between these two appliances. Surprisingly, one device showed high consistency, while the other demonstrated lower consistency. This suggests that the effectiveness of widget-specific gestures is influenced by factors beyond their association with a particular widget.

For state control commands, which involve toggling between different states such as playing, pausing, on, and off, our study revealed low consistency. This lack of consistency was not closely tied to the specific device or scenario. As observed in numerous prior studies, low consistency has been identified for these commands [HTWH19; VK19]. Our analysis indicates that our approach did not significantly contribute to enhancing the consistency of these commands.

Regarding device preferences, participants generally preferred gestures for controlling common smart home devices such as lights, TVs, blinds, windows, heaters, and music players. These gestures mostly involve basic commands like turning devices on or off, opening and closing, and adjusting settings like temperature and volume. One interesting thing we noticed is that dichotomous commands are not equally popular. For example, "next" was a common gesture used 17 times, but "previous" only appeared nine times. This shows that although the gestures suggested for them are often similar but in opposite orientations and have the same level of consistency, people have different preferences for their use. Participants' insights during thinking-aloud sessions revealed that, for personal devices such as smartphones and laptops, proximity is a key factor influencing their choice of interaction modality. They expressed a preference for other interactions, such as touch and mouse, when the devices are in closer proximity. Furthermore, in situations where manual intervention is inherently integral to the operation of a device, such as with coffee machines, participants leaned towards traditional methods for interaction.

### 4.3.2 Scenario Design and User Familiarity

Before beginning the elicitation phase for each scenario, we asked each participant about their familiarity with the priming frame, whether they had encountered it before and to what extent (e.g. if they had ever engaged in a home exercise with a stationary fitness bike). We expected that participants' familiarity with the frame would influence their engagement in each scenario. Among the scenarios examined, participants reported the highest familiarity with cooking, entertainment, and the home office, considering them to be part of their daily routine. Conversely, fitness training emerged as the least integrated scenario into participants' daily routines (only 7 participants reported they ever used a stationary fitness bike for home exercise). Our investigation into mid-air gestures within smart home environments revealed the influence of the participants' familiarity and daily-life integration on the efficacy of our approach. Participants encountered more challenges, such as difficulties in determining which devices and commands are more practical for them to use with mid-air gestures, and exhibited less enthusiasm in less familiar scenarios, such as the fitness training scenario. Hence, our study emphasizes the impact of familiarity with the chosen scenario on user interaction. This may apply to other non-familiar environments, such as a virtual museum exhibit (which is likely to be less familiar to many people). Future studies should investigate this phenomenon.

From a design perspective, effective scenario construction should include:

1. Person's role: This involves defining the specific roles the person undertakes within the environment. For example, in the TV and entertainment scenario, the person's role is illustrated by engaging in an entertainment activity, specifically watching television. In contrast, in the fitness training scenario, the individual's role is characterized by physical exercise.
2. Place and location: This component encompasses the spatial context where the individual is situated and the interaction happens. For instance, in the TV and entertainment scenario, the location is described as a comfortable couch in the living room. Conversely, in the fitness training scenario, the location is conceptualized as the exercise area, in our case, on a stationary fitness bike.
3. Interacting devices: This involves identifying the main devices with which an individual interacts, manipulates, or engages. For example, in the TV and entertainment scenario, the central device for interaction is represented as a television. Conversely, in the fitness training scenario, the primary device is a stationary fitness bike.

However, for the purposes of eliciting devices, commands, and motivations underlying mid-air gesture preferences, we believe that in addition to the central device that must be present, the level of furnishing and other contextual factors that influence ecological validity play a significant role. Participants noted that the realistic environment facilitated a better understanding of their needs and preferences for gesture-based interactions with a smart home. Nevertheless, it is important to note that we did not conduct a comparative study; therefore, careful consideration should be given to the level of detail in layering for future research endeavors.

### 4.3.3 Applicability Beyond Tested Scenarios

The scenarios in our study were chosen for their distinct characteristics, which include variations in physical postures (e.g., standing, sitting, and physical movement) as well as diverse activities and roles undertaken by users within each scenario. We expected that these would help to assess users' preferences in a more diverse range of real-world settings and consider the consistency of users' suggestions across more diverse scenarios. Even though there are inherent differences in each scenario compared to others, similarities between these scenarios and other scenarios in smart homes offer insights that can be extrapolated to other smart home activities. For example, the study's exploration of user preferences and motivations during cooking may serve as a key reference point for understanding interactions in some other kitchen-related activities, such as washing dishes. It is important to note that the scenarios investigated in this study may not encompass all possible scenarios encountered in smart homes. However, we believe the insight from the post-study interviews where participants expressed reservations about mid-air gestures in specific rooms (Section 4.2.7.), contributes valuable insights for researchers designing future scenarios.

### 4.3.4 Comparison to Other Gesture Sets

Comparing our results with other gesture sets developed for smart homes, we see similarity in some cases. For instance, studies conducted by Vogiatzidakis and Koutsabasis [VK19] and Choi et al. [CKL+12], encompassing end-user elicitation studies for different devices in smart homes, exhibited similarities in both gestures and agreement rates for some gestures. For example, the use of "swipe left" and "swipe right" to operate TV controls such as "next" and "previous", as well as "swipe up" and "swipe down" to adjust "music player". Additionally, agreement rates were consistently low for gestures related to commands such as "increase" and "decrease" for controlling volume and temperature in devices like air conditioners and TV, as well as for commands like "turn on" and "turn off".

Guesgen and Kessel [GK12] developed a designer-defined gesture set for home appliances for physically impaired users. They also implemented swipe left and right gestures for navigation commands like "next" and "previous". Hosseini et al. [HIC+23] analyzed 172 relevant papers in the literature on mid-air gestural interaction. They developed a gesture vocabulary consisting of 22 gestures that are highly consistent across different application domains, including smart homes. They also identified different directional swipes for increasing and decreasing, as well as for navigating to the next and previous items. Commands such as "open door", "previous", "next", "zoom", "increase", and "decrease" received high consistency. Their analysis also showed that abstract gestures like "turn on, "turn off, and "play" had low consistency. Our findings are in line with these previous results. This suggests a stable trend towards some gestures becoming standardized also within the context of smart homes.

### 4.3.5 Sensing Technology and the Implementation of Our Findings

Although in our study our initial goal was to explore user-designed gestures without the limitation of current sensing technology, it is crucial to connect our gesture findings with the current sensing technologies and consider the feasibility of implementing our results in real smart home settings.

Our study revealed that participants demonstrated a preference for manual gestures, encompassing movements associated with the hands, forearms, and arms. This type of gesture constituted 91% of all observed gestures, and all of the highest-proposed gestures for various commands, except the sleeping position and move hand away from ear, fell within this category. Additionally, participants displayed a preference for using a single hand to execute commands. Consequently, precise recognition of hand gestures emerges as an essential factor for the success of gestural interaction.

While some of the well-known commercial gesture recognition systems, such as inertial sensor-based and vision-based approaches, have been extensively employed for gesture recognition and can exhibit robust performance, showcasing proficiency in detecting common gestures as outlined in this paper, some limitations exist for both paradigms that hinder easy applicability in our smart homes. Inertial sensors can introduce user discomfort due to the requirement of wearing contacts in diverse smart home scenarios. Conversely, vision-based systems encounter challenges related to the risk of privacy leakage, as well as installation complexity and cost, as we have to deploy dedicated devices and sensors across various locations, which hinders scalable deployment, especially in our case where we address gestures across entire homes and different rooms. Therefore, we recommend the adoption of Wi-Fi or radar-based techniques for entire home gesture interaction as suggested in previous studies [PGGP13; WLLP14]. Some studies have specifically investigated the feasibility of using Wi-Fi or radar-based techniques by designing prototype implementations with off-the-shelf devices. For instance, Nandakumar et al. [NKG14] investigated the feasibility of using Wi-Fi off-the-shelf devices in their gesture recognition system, achieving 91% accuracy. The gesture set included different hand gestures like push, pull, lever, and punch. In another study, the focus was on radar-based gesture recognition, where a low-complexity algorithm using frequency-modulated continuous-wave (FMCW) radar was proposed. This algorithm achieved an average recognition accuracy of 89.13% for a gesture set, including grab and swipe gestures, among others [ZSKG23]. As a result, the demonstrated capability of these technologies to recognize a variety of gestures underscores the suitability of these technologies for the gesture set described in this paper.

In our study, employing similar gestures for identical commands across multiple devices, while maintaining a concise gesture set, is beneficial for memorability. Nevertheless, this approach poses challenges, especially in terms of implementation and designing such systems. To ensure seamless implementation in such a multi-device environment, addressing the Midas touch problem is essential. This problem especially arises when user interaction with one device unintentionally activates others, due to similar gestures. There are several methods proposed for addressing this challenge. Addressing and commands are the primary approaches suggested for multi-device interactions, where users first address a specific device and then provide a command. Examples include pointing and commands [VK22]. For implementation, vision-based recognition (using cameras) for each device is often employed. This allows devices to recognize gestures, ensuring that a device understands when a performed gesture is directed towards it. However, this approach raises the issue mentioned earlier, particularly in scenarios where users may be in different rooms; therefore, pointing to the device is not effective. However, we believe it is feasible to implement our resulting gesture set using addressing and command methods. For addressing, we propose the use of gestures, such as pointing to a specific part of the body (e.g., pointing to the eyes to address the TV) or indicating a universal infrastructure in a building, such as pointing to a cell, for addressing light as suggested by [VK22].

### 4.3.6 Design Implications

In this section, we discuss the implications of our findings regarding the design of mid-air gestures. We aim to complement other implementations recommended in prior work by offering our new viewpoint on gestural interaction in the smart home.

#### Engaging with Real-World Situations in User Elicitation

In our study, we conducted a scenario-based elicitation investigation within a realistic smart home laboratory environment. By immersing participants in four scenarios: cooking, TV and entertainment, home office, and fitness training, we extracted valuable insights regarding the design of mid-air gestures for interacting with various available devices. This approach allowed us to explore user preferences associated with gesture-based control in real-world smart home settings and maximize the ecological validity of user-proposed gestures. Our study was well received by the participants. They noted that the immersive environment made understanding their own needs and wishes towards gesture-based interactions with smart home devices easier. Additionally, participants found the study to be more engaging. Consequently, expanding upon prior research [VK19; CLA18], we propose that elicitation studies incorporate contextual elements to enhance the ecological validity and the intuitiveness of the suggested gestures.

#### Consistency in User Commands Across Devices for Universal Gesture Patterns

Our user study revealed consistency among participants across various devices for certain frequently mentioned commands. Commands such as "open" or "close" for door or window control, "volume up" or "volume down" for music player manipulation, and adjustments in temperature settings ("temperature down" for heaters and stoves, "temperature up" for heating devices) had a high level of uniformity. It is thus evident that users tend to associate specific gestures with controlling different devices. Therefore, in the context of commonly selected commands that display a significant degree of consistency across diverse devices, we advise designers to prioritize the development of universal gesture patterns.

#### Gesture Interaction and User Confidence in Sensitive Tasks

During the study, some participants expressed reservations about using gestures because they were concerned that unintentional movements might trigger the system to perform actions, especially in potentially hazardous contexts such as turning on an unsupervised oven or stove. To address this issue, we recommend implementing feedback mechanisms to increase user confidence when employing gesture-based interactions in such situations.

Providing clear visual or auditory feedback can help users feel more in control and reduce the likelihood of accidental activation.

### Exploring Fun and Intuitive Gestural Commands

While the primary motivations of users for integrating gestural interactions often include comfort, hygiene, reduced interruptions, and time savings, it is noteworthy that some participants in our study proposed gestural commands because they found them ‘fun’ to perform, enhancing the overall enjoyment of the interaction. For instance, gestures like clapping hands or snapping fingers were chosen by participants because they believed these actions could inject an element of enjoyment into the interaction. Therefore, these types of gestures, along with other gestures identified in the literature, such as body-referenced gestures [Vat13c; KWH+11], serve as valuable resources for interaction designers seeking to infuse fun into their interactions.

### Limitations and Future Work

We would like to highlight that all our participants were from Central Europe. Due to this specificity, we were able to thoroughly examine gesture communication within the context of this particular culture. We did not examine the impact of cultural factors on gesture meaning, despite well-documented differences in gesture meaning across cultures [MHWC10; RM17], which limits the generalizability of our findings.

Additionally, while participants provided valuable insights into their preferences for mid-air gestural interactions in a smart home environment, one of our limitations is the lack of empirical validation. In future work, we aim to address this limitation by conducting comprehensive empirical validation studies to evaluate the usability, efficiency, and user satisfaction associated with these gestures in real-world scenarios.

Furthermore, in this study, participants were given the freedom to decide on both the device and command, resulting in an uneven distribution of proposed gestures. Therefore, it is important to note that while some commands (such as “speed up” and “move between pages”) achieved high agreement, the overall number of gestures proposed for these commands remains limited. This limitation may impact the generalizability of less frequently mentioned commands. We recommend that future researchers consider this when applying insights from these commands.

## 4.4 Conclusion

This chapter presented a scenario-based elicitation study focused on mid-air gestural interaction within smart homes. We explored people’s attitudes toward adopting mid-air gestures across various smart home scenarios and for diverse devices. Our investigation involved measuring agreement rates across devices and scenarios to shed light on the possibility of establishing a consistent and universal gesture set to simplify user experiences by eliminating the need to remember different gestures for similar commands applicable to multiple devices within the home. Our research revealed that people’s motivations for adopting mid-air gestures vary across different home scenarios. Scenarios related to television and entertainment place a high priority on comfort, while those related to cooking place a high priority on hygiene. Our analysis of user-defined gesture sets revealed variations in agreement levels, ranging from very high to low. In addition, our findings suggest that most smart home users prefer consistent gestures. Also, some gestures demonstrated high consistency between commands applicable to multiple devices. Our approach immersed participants in real-life scenarios and yielded valuable insights and recommendations closely aligned with users' needs. Future work stemming from this paper will design and implement these gestures within our smart home laboratory environment. We intend to empirically evaluate their usability and overall effectiveness.

# 5 Development of a Multimodal Mid-Air Gesture Dataset

The study of mid-air gestures in interactive systems requires not only identifying which gestures users prefer, but also gaining detailed insight into how gestures are physically executed. We examined user-defined gesture vocabulary and their consistency across scenarios and domains in the previous chapters. While these studies provide insight into intuitive gesture–action mappings, they often capture only descriptive information about gestures or static visual representations of intended gestures, rather than details about the way in which those gestures are actually performed in practice. Addressing this limitation requires a high quality dataset that captures the detailed motion characteristics of gesture execution across users.

Although many gesture and motion datasets already exist, they tend to be task-specific, domain-limited, or sensor-restricted. Many existing motion datasets often focus on full-body activities such as dance, sports, or everyday actions (e.g., NTU RGB+D [XCW+19], UCF50 [RS13b], Actions as Space-Time Shapes [BGS+05b], Actions Dataset [SLC04]). Other datasets concentrate on specific hand gestures, such as grasping or sign language (e.g., GrabMyo [JPH22], ObMan [HVT+19], Ninapro [PCG+17]). Many existing resources also rely on constrained recording setups or limited sensing modalities, such as RGB video (e.g., HMDB51 [KJG+11]), accelerometer-based sensing (e.g., WISDM [KWM11]), or electromyography (EMG) (e.g., UC2018 [SNG18]). Collecting multimodal data is challenging, as adding more sensors increases the complexity of cross-device synchronization, calibration, and data management. As a result, multimodal datasets with fine-grained spatiotemporal tracking remain rare or typically involve a limited number of participants. As a result, there remains a lack of high-fidelity, multimodal datasets that comprehensively capture a broad set of mid-air gestures drawn from diverse application domains and interaction contexts. Such datasets are essential for analyzing how gestures are executed across different users and repetitions, which is the focus of our work.

To the best of our knowledge, no publicly available dataset covers the full set of gestures examined in this thesis, which have been identified as the most frequently proposed gestures in interactive systems across diverse domains. Additionally, they do not necessarily provide a comprehensive capture of these gestures using synchronized multi-view RGB-D recordings together with high-fidelity positional time-series data. To address these limitations, we developed a novel mid-air gesture dataset titled MAGIC, comprising a diverse set of gestures frequently identified in interactive systems. The gesture set we captured was derived from our literature review (Chapter 3) and elicitation study (Chapter 4) results. The dataset comprises 18 commonly proposed uni-manual mid-air gestures with the potential to be considered transferable across domains. Each gesture performed by 42 participants across multiple trials. This results in a corpus of 4,265 gesture instances. It integrates three synchronized sensing modalities to capture both fine-grained hand motions and global full-body posture while gestures are being performed. Specifically, we employed a 14-sensor electromagnetic tracking system to record high-resolution 3D positional time-series of the hand and arm, enabling analysis of upper-limb kinematics. In parallel, three depth cameras were employed to capture the full-body pose and movement during the execution of each gesture, yielding RGB and depth video data recordings.

In this Chapter, Section 5.1 describes the study design, including participant demographics, sensing apparatus, experimental procedure, and ethical considerations. We then describe the structure and contents of the collected dataset, including preprocessing steps and data formats and content, in Section 5.2. Finally, Section 5.3 discusses the dataset and its limitations.

The dataset described in this chapter has been accepted for publication as part of a dataset paper in the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) by Hosseini et al.

## 5.1 Study

This section presents the design and implementation of a data collection study conducted in a controlled lab environment. The study generated a multimodal dataset of mid-air hand gestures by using depth camera and electromagnetic sensor technologies. This dataset enables us to investigate the performance of gesture execution and memory recall in the following Chapter.

### 5.1.1 Participants

A total of 42 participants (30 female, 12 male) aged 18–38 years (median = 25 years, IQR = 23–29 years) participated in our study. Participants primarily had bachelor's (n = 27) or master's (n = 8) level educational backgrounds, while some had secondary (n = 7) level education. They were recruited through the university's online notice board and research group environment. The study lasted approximately two hours, and participants were reimbursed for their time.

All participants confirmed, to the best of their knowledge, that they were free from any musculoskeletal disorders or anatomical features that would limit the motor ability of their hands and fingers. When asked about hand dominance, 40 participants reported being right-handed, and two stated that they were ambidextrous. As instructed, all participants used their right hand for the gestures.

Participants rated various aspects of gesture-based interaction using a scale from 1 to 5. They gave an average rating of M = 2.17 (SD = 1.36) for their comfort using hand or body gestures in their living environment. The average score for the frequency of hand or body gestures within their cultural context was M = 2.07 (SD = 0.89). Participants also reported an average of M = 1.88 (SD = 1.06) hours per week spent engaging in repetitive upper-limb physical activity, such as sports, musical instrument practice, or manual tasks. Most participants indicated little to no prior experience with gesture-based control. A number of participants (n = 9) reported occasional interaction with systems such as Xbox Kinect, PlayStation EyeToy, VR sports games, or research studies involving mid-air gestures.

### 5.1.2 Apparatus and Setup

The recording took place in a room designed specifically for this purpose, which minimized environmental noise and ensured consistent recording conditions for all participants. To reduce visual distractions and minimize background artifacts in the video data, the walls were uniformly covered with white fabric, thereby enhancing the clarity of participants' gestures. Fixed, white overhead lighting was used throughout each session. Within this environment, we set up a multimodal sensor system to record participants' hand movements. Specifically, we employed three Intel RealSense D455 cameras [Int] and a Polhemus Viper16 electromagnetic tracking system [Pol].

As illustrated in Figure 5.1(a), three Intel RealSense D455 cameras were positioned around the participant to capture a wide range of hand movements and overall body posture. One camera was placed on the front left side (C1), another (C2) was positioned at the front right. Additional camera was located at the back right (C3). The cameras were configured to capture color streams at 30 frames per second (fps) and depth streams at 90 fps. This configuration allows for a broad field of view but with relatively lower resolution compared to the electromagnetic tracking system.

To achieve higher-resolution tracking, we utilized the Polhemus Viper16 electromagnetic tracking system, which employs 16 tethered microsensors to track the participant's right hand and fingers at 960 fps. Five sensors were attached to the fingertips, and five sensors were placed at the middle of the proximal phalanges of the fingers. Additionally, one sensor was attached to the metacarpal of the middle finger, one to the head of the radius, one to the end of the humerus, and one to the shoulder. Two additional sensors were integrated into the hand rest position on the right side of the participant, enabling the detection of the onset and termination of hand movements. For Viper tracking in this study, we used the raw 3D coordinates provided by the Viper electromagnetic tracker, with the origin at the position of the electromagnetic field source and the axes aligned with the orientation markings on the source unit, where the x-axis points up.

All devices were synchronized to ensure precise temporal alignment across data streams. Following the hardware synchronization procedure described by Grunnet-Jepsen et al. from Intel [GJWT+18] and software implementation at [Val25], three RealSense D455 units were connected via hardware synchronization cables, with the depth sensor of the front-left camera configured as the master device and the remaining cameras as slaves. Master camera created a sync signal at the initiation of each depth-frame capture cycle. This signal was distributed through the hardware sync cables to the remaining cameras and was also used as an external timing reference by the electromagnetic tracker. The tracker recorded these trigger transitions at its internal sampling rate, producing digital markers that indicate the onset of each camera frame within the motion-tracking data stream.

The data capture application was developed using C/C++ and ran on a Windows 11 laptop equipped with an Intel Core i9-14900HX processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4070 graphics processing unit (GPU).

### 5.1.3 Procedure

Before the study began, the participants received an overview of the study and its experimental setting. Then, they were presented with the informed consent document for reading and signing. After that, participants were positioned in the recording area. The study consists of four phases: Calibration, Recording with Instruction, Recording without Instruction, and Post-Study Questionnaire. To provide standardized instructions between

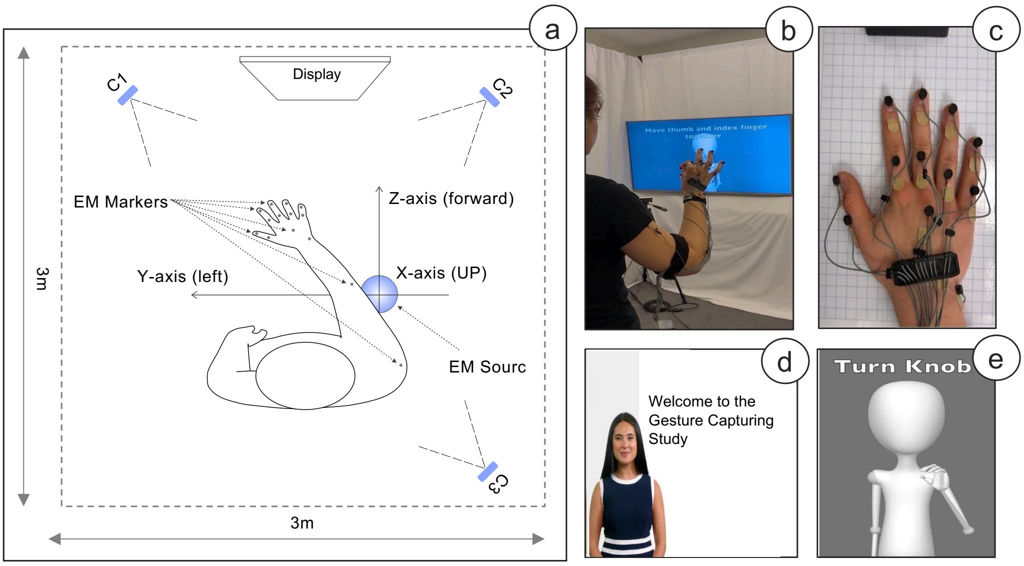



**Figure 5.1:** (a) Experimental setup showing the participant's location, surrounded by three depth cameras: C1 (front left), C2 (front right), C3 (back right). The 3D coordinates provided by the VIPER tracker, with the origin at the source, and where the x-axis points upward. (b) A participant performing hand gestures with sensors attached to the hand and fingers. (c) Sensor placement on the participant's hand and fingers. (d) An AI-generated instructional video was used to guide the participant. (e) The avatar used for animating gestures.

participants, we developed an AI-generated video using synthesia[1] that presented consistent information about the study procedures to ensure consistency among participants and improved the study's reproducibility.

The study consisted of the following phases:

1. **Calibration Phase:** During this phase of the study, magnetic sensors were attached to participants' right hands at the locations described in Section 5.1.2. Then, hand measurements were taken using a top-down document camera while the hand was positioned on graph paper. Participants were then instructed to place their hand and arm flat on the resting surface of the hand to establish a consistent baseline for tracking that could serve as a reference.
2. **Recording with Instruction Phase:** After the calibration phase, the recording with the instruction phase began. For each gesture in our set, the participant was shown an instructional video of a genderless character (see Figure 5.1(e)) performing the gesture [KAC+18]. These videos were created using Blender, based on a predefined set of gestures animated by the experimenters. After each gesture was shown, the participant was asked to perform it. After each execution, the participant rated their confidence in their execution of the gesture on a scale from 1 (not confident) to 10 (very confident). Each participant performed each gesture three times in a randomized order. This phase was designed to ensure a consistent understanding of the gestures among participants and to establish a baseline for performance under guided conditions.

[1] synthesia, `https://www.synthesia.io/`

3. **Recording without Instruction Phase:** In this phase, participants were asked to recall and perform the same gestures they had performed in the Recording with Instruction phase, but without instructional assistance. Participants were provided only with the name of each gesture in the gesture set. If they remembered a gesture, they were asked to perform it. After each performance, participants rated their recall of the gesture and their confidence that their performance matched the intended gesture on a scale from 1 (not confident) to 10 (very confident). Each participant performed each gesture three times in a randomized order. This phase aimed to capture more natural gesture executions in the absence of visual guidance and to evaluate the memorability of gestures after initial exposure.
4. **Post-Study Questionnaire:** Afterward, each participant was asked to answer questions about their demographics and provide feedback about their experience using mid-air gestures.

Breaks were taken between phases or when participants felt exhausted. Our study employed a within-subjects design and was conducted in single-user sessions.

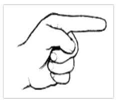
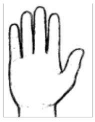
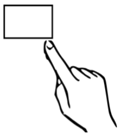
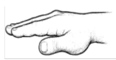
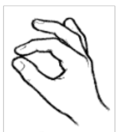
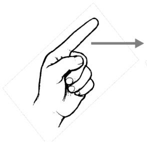
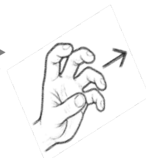
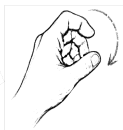
### 5.1.4 Selected Gestures

The selection of gestures for recording was based on our systematic literature review of 172 studies [HIC+23], which extracted 1,728 mid-air gestures across diverse interactive system domains. These gestures were grouped into thematic categories, such as pointing, swiping, and scaling. Each category represents a gesture type with varied physical executions (e.g., pointing with the index finger vs. pointing with an open hand or horizontal vs. vertical swiping). These gesture types reflect recurring patterns commonly observed across different domains. For recording, we focused on the 18 most prevalent gesture types identified in this review. From each type, we selected the most frequently reported unimanual variant. This process yielded a set of 18 distinct uni-manual gestures intended to reflect commonly used forms of mid-air gestures. Table 5.1 presents the final set, which includes 11 dynamic and seven static gestures.

### 5.1.5 Ethical and Data Anonymization

The Ethics Committee of our institution gave ethical approval to the study, and informed consent was obtained from all participants. Since the RGB videos contained identifiable traits of the subjects, we anonymized the video sequences by applying face pixelation using an automated OpenCV and MediaPipe. This detects face regions and replaces them with coarse pixel blocks to remove identifiable information, while protecting hand regions from corruption, thereby preserving the dataset's utility for gesture recognition research. Each video was reviewed to confirm the quality of anonymization, and any cases where the automatic procedure failed to pixelate faces were manually anonymized. The depth frames were left unmodified because they do not contain identifiable information, as our setup has low resolution due to the sensors being far away and partial face occlusion caused by the sensor's positioning, which prevents reliable facial reconstruction, with current AI methods [Jun20].

**Table 5.1:** List of the 18 motion gestures captured in the dataset.

| | | | | | |
|---|---|---|---|---|---|
| 1- Point right | 2- Show palm | 3- Thumb up | 4- Show number 3 | 5- Show fist | 6- Tap with index |
|  |  |  |  |  |  |
| 7- Double tap with index | 8- X symbol | 9- Draw a rectangle | 10- Lingering open hand | 11- Move thumb and index together | 12- Move apart thumb and index |
|  |  |  |  |  |  |
| 13- OK sign | 14- Point and move | 15- Grab hand and move forward | 16- Turn Knob | 17- Wave hand | 18- Swipe hand |
|  |  |  |  |  |  |

## 5.2 Results

### 5.2.1 MAGIC Dataset

This section provides the main information about the dataset's structure, content, and the preprocessing procedures applied. The dataset is publicly available at: https://doi.org/10.57782/ACBGC8.

#### Data Pre-processing

To ensure the quality and consistency of the collected gesture dataset, a post-acquisition cleaning and validation process was performed. All color stream recordings were manually reviewed to identify and remove incomplete or low-quality samples. Specifically, gestures involving visual artifacts, such as motion blur, occlusion, or sensor detachment, were excluded from the analysis. In addition to visual clarity, the semantic correctness of each gesture was validated by comparing it with the example gestures provided to participants. Trials in which participants performed a completely different gesture (e.g., showing a thumb up instead of making an okay sign) were excluded. However, we observed instances in which participants deviated slightly from the instructed gestures. For example, they performed three taps instead of a double tap. These minor deviations were retained in the dataset because we considered them representative of the natural variation in how participants perform gestures. This aligns with the aim of our study, which is to examine how participants naturally perform gestures. Additionally, off-task behavior, such as random movements during a gesture or irrelevant hand movements, was excluded. Furthermore, all positional time-series data from the Viper tracking system were evaluated for sudden jumps, incomplete frames, invalid marker indices, high distortion (above 50), and non-numeric or incomplete positional values. Frames with fewer than 16 detected markers were retained; however, the missing sensor values were marked as NaN (not a number) in the dataset to indicate unavailable data and maintain data continuity. Frames with distortion levels exceeding 50 were removed, as higher distortion values are associated with unstable tracking and increased positional jitter. All data preprocessing steps were carried out using custom Python scripts. Moreover, for each gesture, annotations were validated by reviewing the recorded video and motion data to verify that they accurately corresponded to the action performed. Additionally, temporal alignment consistency was verified to ensure that all modalities shared the same start and end frames and had consistent trial durations.

Furthermore, during data preprocessing, slight misalignments were identified in the fingertip sensors of the ring and index fingers. To correct these misalignments, we used the middle finger, which typically exhibits the most stable and consistent movement, as a reference. For gestures where the fingers move in coordination (such as showing the palm or waving), the index, middle, and ring fingers tend to move together while maintaining a stable relative shape. We first calculated the displacement of the middle fingertip sensor relative to its proximal sensor at each time frame to capture the finger's movement. This was then scaled according to the length of each finger (index and ring) and added to the position of their respective proximal sensors to estimate the corrected fingertip positions. In contrast, for gestures in which the index finger moves independently (e.g., pointing or tapping), its motion was estimated using the direction and magnitude of movement from its proximal sensor, scaled to reflect typical finger curvature and extension. Corrections were applied only when misalignments were detected, and all adjustments were validated through visual inspection of the plotted sensor trajectories and video recordings to ensure the realism of the gestures (See Figure 5.2).

While low-quality or incomplete trials were removed during Preprocessing, two participants were recorded under slightly different ambient lighting conditions within the same recording environment. These recordings were retained because the positional time-series data met semantic and temporal validation criteria. We flagged these sessions in the metadata so that researchers can easily identify them.

### Dataset Structure and Contents

A total of 4,265 motion sequences were retained in the final dataset from the 42 participants who took part in the experiment, along with a questionnaire capturing participant demographics and a calibration folder. The total dataset size is 1.25 TB.

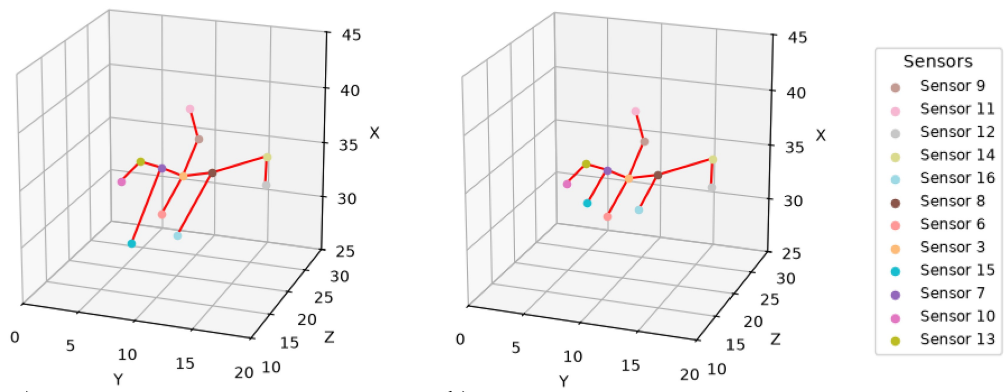



**Figure 5.2:** Visualized sensor plots of fingers. a) misaligned positions of Sensors 15 and 16, b) corrected sensor positions after applying correction.

The dataset comprises three synchronized sensor modalities: RGB video, depth video, and position data. Recordings are grouped into anonymized participant folders, each containing a calibration subfolder and labeled gesture trials. Each gesture trial is named according to the structure `<Gesture>_<TrialID>_<Condition>`, indicating gesture label, repetition number, and study phase.

Each recording gesture sample (trial) contains 13 data files, including color and depth video from three RealSense D455 cameras (encoded using the UT Video codec), per-frame metadata text files for each stream, and a CSV file containing Viper sensor readings for that gesture sample. Depth values are encoded across red and green color channels. Viper data includes 3D positions (in centimeters) and binary button states (hardware sync signal) for all 16 sensors sampled at 960Hz. The calibration folder includes chessboard-based recordings (Record grid), as well as record hand, record setup sessions, and participant demographics.

## 5.3 Discussion

In this chapter, we introduced a novel multimodal dataset designed to support research in mid-air gesture interaction. Two sensing modalities were employed in the dataset: (1) a depth camera that captures both RGB and depth video data, and (2) wearable electromagnetic sensors that provide time-series motion data of the hands and arms. Unlike existing gesture datasets that often focus on task-specific functional movements (e.g., sign language or grasping) or are constrained by limited sensing configurations, our dataset, titled MAGIC, was constructed using a combination of high-fidelity motion capture systems and commodity depth sensors. It captures diverse communicative gestures drawn from human-computer interaction scenarios and performed by 42 participants under recall and learning conditions. This dual-modality approach enables fine-grained analysis of gesture execution characteristics such as kinematics, reproducibility, and inter-user variability.

With MAGIC, we provide both the interaction design and multimodal motion analysis communities with a domain-relevant benchmark dataset to support the development and evaluation of robust mid-air gesture recognition and interaction techniques.

### Limitations and Future Work

Although our data collection setup enables multi-modal gesture capture, several limitations must be acknowledged. First, variability among participants may affect the generalizability of the dataset. The study included young, non-disabled individuals; thus, it does not accurately represent diverse user populations, such as the elderly, people with limited mobility, or individuals from different cultural backgrounds. These groups may exhibit distinct gesture styles. Additionally, although some left-handed individuals participated, all gestures were performed using the right hand, which may limit insights into behavior for left-handed users.

Second, the hand-mounted sensors used for gesture capturing may restrict the natural range of motion or affect the spontaneity of certain gestures. Although care was taken to minimize encumbrance, the placement of the sensors could subtly alter the kinematics of the gestures. Third, although the depth cameras' resolution is sufficient for most recognition tasks, it requires a trade-off between spatial precision and dataset size. While higher resolution could improve motion analysis, it would result in larger files and potentially greater privacy concerns. In addition to addressing the identified limitations, future work should include multi-environment recordings to evaluate performance under different lighting conditions, levels of background clutter, different sensor placements, and other environmental factors. .

# 6 Analysis of Gesture Execution and Performance

As mid-air gesture interfaces become more prevalent in state-of-the-art systems such as Microsoft Kinect [SFC+11], Leap Motion [WBRF13], and AR/VR platforms such as Microsoft HoloLens [BCL+15], it is critical to design systems that accurately interpret user motion, support expressive and diverse gestures, and remain robust across users and environments. Achieving this requires an understanding of how gestures are physically executed by users in practice, such as their motion patterns, timing, articulation consistency, and variability across individuals.

Previous research has introduced numerous algorithms and tools for gesture recognition using methods such as template matching [RYZ11], dictionary look-up [COPB12], statistical matching [New94], neural network [KK12], and ad hoc method [MDZ16]. Complementary to this, other studies have focused on the semantic definition of gestures using user-driven elicitation methods. For example, Wobbrock's "Guessability Study" [WMW09], Choice-based elicitation Study [DSSR16], Nielsen's "Intuitive and Ergonomic method" [JNCR18], have contributed valuable insight into gesture vocabulary design across contexts.

However, while these studies reveal user preferences and contribute to the development of intuitive gesture-action mappings, they usually only capture users' intended gestures rather than the users' actual performance behavior and ergonomic aspects such as fatigue, comfort, and ease of performance, which are critical for evaluating whether proposed gestures are physically feasible and reproducible across users [SSC23; VK18]. To address this limitation, this thesis takes a complementary approach by analyzing what users actually do. Using a dataset of motion-tracked gesture performances that we collected, we evaluate the physical execution of commonly proposed gestures in interactive systems across diverse application domains.

In this chapter, we analyze the mid-air gesture dataset collected and described in Chapter 5 in order to better understand how individuals naturally perform gestures and to identify the variability and commonalities in gesture execution. To achieve this, we analyze our dataset using a combination of general features, such as gesture execution duration or movement range, along with other quantitative measures proposed in prior work [Vat17]. We also employ Multidimensional Dynamic Time Warping (MDTW) [Mü07] to compare the multi-sensor 3D positional data of hand gestures captured using electromagnetic motion tracking, enabling the analysis of hand articulation patterns and the assessment of their temporal similarity [LMAV17]. Specifically, we assess intrapersonal variability of gestures by comparing multiple repetitions of the same gesture performed by a single individual, and interpersonal variability by analyzing how different individuals perform the same gesture. Additionally, we compare different gesture types to evaluate their distinctiveness and consider their applicability in real-world interaction scenarios. To evaluate the suitability of our dataset for gesture recognition, we also employ a Long Short-Term Memory (LSTM)-based recognition model as a baseline under both user-dependent and user-independent settings. Parts of this chapter have been accepted as a full paper in the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) by Hosseini et al. together with the dataset described in Chapter 5.

## 6.1 Measures for Gesture Analysis

To analyze the recorded gestures and characterize their variability and consistency across contexts, we extracted a set of geometric and kinematic features, adapted from prior work on stroke-gesture recognition and Laban Movement Analysis [Vat17; AVW13; SA16]. We additionally used MDTW to evaluate gesture similarity across repetitions. We further trained LSTM-based classifiers to establish recognition baselines and compute a deep-feature consistency measure from the learned embedding space. All features were computed from the time-series data collected using the EM tracker. We define a gesture $G$ as a sequence of 3D positions recorded from $S = 14$ upper-limb sensors.

$$G = \left\{ P_t \;\middle|\; P_t = \{p_{t,s} \mid p_{t,s} = (x_{t,s}, y_{t,s}, z_{t,s})\}_{s=1}^{S} \right\}_{t=1}^{n} \tag{6.1}$$

Here, $P_t$ represents the set of 3D positions of all sensors at time step $t$, and $p_{t,s}$ denotes the 3D position of sensor $s$ at time $t$. Each $p_{t,s}$ is defined by its spatial coordinates $(x_{t,s}, y_{t,s}, z_{t,s})$ in a world-space coordinate system defined by the Viper electromagnetic tracker.

### 6.1.1 Geometric and Kinematic Features of Gesture Execution

We extracted geometric and kinematic features to describe gesture execution. Geometric features capture the spatial characteristics of gesture execution. We analyze three features: quantity of gesture movement (QM), gesture volume (GV), and gesture area (GA). These features have previously been used to characterize gesture movement and have been associated as indirect proxies for physical demand, potential fatigue, and social acceptability [HRGMI14; EQY+18]. Larger gestures may indicate higher physical demand and potential fatigue [HRGMI14; EQY+18], serving as objective proxies for physical demand in the absence of direct measurements, and may also be less socially acceptable in public settings [AHI14]. Furthermore, they can also serve to inform the design of user interfaces employing hand gestures, for example, by determining the volume of space around the upper body required for users to perform gestures safely, in line with recommendations and operating guidelines for hand motion capture systems [VGO+20].

Kinematic features, in contrast, describe the temporal and dynamic aspects of gesture performance. We analyze two kinematic features: performance time (PT) and speed (S). These features have been used to assess user fluency and memorability. Faster execution often indicates greater familiarity or ease of recall, whereas slower or more variable performance can suggest difficulty in execution or challenges for system interpretation [CAMCD+19].

#### Geometric Measures

**1.** *Quantity of Gesture Movement*, represents the overall upper-limb displacement during gesture execution. It is computed by summing the Euclidean distances between consecutive time frames across all relevant sensors along the upper limb:

$$QM = \frac{1}{\lambda} \sum_{t=2}^{n} \sum_{s=1}^{S} \left( \| p_{t,s} - p_{t-1,s} \| \right) \tag{6.2}$$

$\|p_{t,s} - p_{t-1,s}\|$ represents the Euclidean distance between consecutive positions of the same sensor $s$, for $s = 1, \ldots, 14$ and $t = 2, ..., n$. The parameter $\lambda$ is a normalization factor determined by the application context. We set $\lambda = S = 14$ such that QM reflects the average displacement per upper-limb sensor (excluding the two hand-rest sensors used for onset/offset detection). To reduce jitter and sensor noise in the raw trajectory data, we applied a smoothing process prior to computing the QM. First, each 3D sensor position was passed through a median filter with a kernel size = 5. Then, a first-order infinite impulse response (IIR) low-pass filter with coefficient $\alpha = 0.1$ was applied to each dimension. This smoothing process was also applied to calculate the other measures in this section. Since QM averages displacement across all 14 sensors, the contribution of fine-grained finger movements may be diluted by larger proximal sensor displacements; finger configuration and hand shape are instead characterized by the consistency measures in Section 6.1.2.

**2.** *Gesture Volume*, represents the 3D spatial extent traversed by the hand during a gesture. It is computed as he volume of the axis-aligned bounding box defined by the hand trajectory, i.e., the product of the differences between the hand's maximum and minimum positions along the x, y, and z axes.

We excluded the arm and elbow from the GV computation. Therefore, this metric reflects on the 3D interaction workspace occupied by the end-effector level (hand), capturing interface layout constraints and spatial requirements of interaction by hand.

$$GV = \prod_{\delta \in \{x,y,z\}} \left( \max_{t,s} \left(P_{t,s}^{\delta}\right) - \min_{t,s} \left(P_{t,s}^{\delta}\right) \right)$$

where $\delta$ represents the dimensions $x$, $y$, and $z$. $P_{t,s}^{\delta}$ is the position of the $s$-th sensor in the $\delta$-th direction for the $t$-th time stamp.

**3.** *Gesture Area,* represents 2D projected area occupied by the hand trajectory within a selected reference plane. The selected 2D reference plane for GA computation depends on the interaction context. For example, in screen-based gestures, GA can be measured in the XY plane (frontal view), while for gesture-controlled drones, GA can be computed in the XZ plane (top-down view). Similar to Gesture Volume (GV), the elbow and upper arm are excluded from GA computation in this paper. The measure therefore reflects the visible or projected footprint of the gesture at the interaction point (end-effector), rather than full upper-limb motion.

$$GA = \prod_{\delta \in \{x,y\}} \left( \max_{t,s} \left(P_{t,s}^{\delta}\right) - \min_{t,s} \left(P_{t,s}^{\delta}\right) \right) \tag{6.3}$$

$\delta$ represents the $x$ and $y$ dimensions (for 2D space). $P_{t,s}^{\delta}$ is the position of the $s$-th sensor in the $\delta$-th direction for the $t$-th time stamp. GA was computed in the frontal plane.

### Kinematic Measures

**1.** *Performance Time,* measures total gesture duration. It is calculated as the difference between the time stamp of the gesture's starting frame and its ending frame.

$$PT = (t_n) - (t_1) \tag{6.4}$$

where $t_n$ and $t_1$ represent gesture onset and offset, respectively.

**2.** *Gesture Speed*, quantifies the mean velocity at which a gesture is performed. It is computed as the ratio of QM (the average per-sensor displacement across all 14 upper-limb sensors) to the gesture performance time (PT). To ensure measurement reliability, position data were preprocessed using the previously described two-stage smoothing. Because QM averages displacement across all 14 sensors, S is not directly comparable across gestures that differ in their degree of arm versus finger involvement (see Section **??**).

$$S = \frac{QM}{PT} \tag{6.5}$$

### 6.1.2 Gesture Consistency Using MDTW

To examine the consistency of gesture performance across gesture trials, we computed pairwise dissimilarities using multidimensional dynamic time warping (MDTW). This function identifies the optimal alignment between sequences by non-linearly warping the time series. Rather than using raw 3D positions, which are sensitive to body size and global position, we employed two feature representations: (i) hand shape, capturing the relative configuration of the hand and fingers, and (ii) palm trajectory, capturing the spatial motion path of the hand.

Due to natural variations in execution speed, trials contain different numbers of frames, making direct frame-by-frame comparisons unsuitable. For comparison gesture trials, each gesture was resampled to a fixed length (N=150), using linear interpolation based on cumulative arc length, following the approach in gesture recognition (e.g. [VU22]).

$$S = \frac{L_G}{(N-1)} \tag{6.6}$$

where $L_G$ is the total arc-length of gesture $G$, computed as the sum of Euclidean distances between consecutive 3D points. New points were inserted along the trajectory at arc-length intervals of size $S$ using linear interpolation between the two original points that surround each target location. This method produces gesture sequences of equal length while preserving their relative motion structure.

**Per-Frame hand-shape via palm-to-fingertip vectors.** At each frame on the re-sampled data, we compute vectors from the palm sensor to each fingertip. Then, to eliminate invariance to hand size and absolute positions, we normalize each vector to unit-length direction vectors. Indeed, if $\vec{f_i}(t)$ represents the 3D position of fingertip $i$ and $\vec{p}(t)$ represents the palm position at time $t$. Then, the direction vector is:

$$\vec{v}_i(t) = \frac{\vec{f_i}(t) - \vec{p}(t)}{\|\vec{f_i}(t) - \vec{p}(t)\|} \tag{6.7}$$

Hand-shape similarity was computed using MDTW, with per-frame cost defined by cosine dissimilarity between corresponding direction vectors. For any two frames, $A$ and $B$, the dissimilarity cost for vector was defined as the average angular deviation across the five fingers:

$$\text{frame\_cost}(t) = \frac{1}{5}\sum_{i=1}^{5}\left(1 - \vec{v}_i^{(A)}(t) \cdot \vec{v}_i^{(B)}(t)\right) \tag{6.8}$$

Here, $\vec{v}_i^{(A)}(t)$ and $\vec{v}_i^{(B)}(t)$ denote the unit vectors from the palm to fingertip $i$ at frame $t$ in gestures $A$ and $B$, respectively. This captures hand pose similarity by comparing relative finger orientations while being invariant to absolute position or hand size.

**Palm trajectory representation.** To capture the spatial motion component of gesture execution, we compute a palm trajectory consistency measure based on the 3D path traced by the palm sensor. Each palm trajectory is first translated relative to the mean 3D position of the hand-rest reference sensor, to remove absolute positional differences across executions, and then scaled by its total path length to emphasize trajectory shape over movement amplitude:

$$\mathbf{p}'(t) = \frac{\mathbf{p}(t) - \bar{\mathbf{s}}}{L}$$

where $\mathbf{p}(t)$ denotes the 3D position of the palm sensor at time step $t$, $\bar{\mathbf{s}}$ the mean surface reference sensor position, and $L$ the total trajectory length.

Gesture similarity under the trajectory representation is then computed using MDTW with a per-frame cost defined as the Euclidean distance between corresponding 3D normalized palm positions:

$$\text{cost}_{\text{traj}}(t) = \|\mathbf{p}'^{(A)}(t) - \mathbf{p}'^{(B)}(t)\|$$

where $\mathbf{p}'^{(A)}(t)$ and $\mathbf{p}'^{(B)}(t)$ denote the normalized palm positions at time step $t$ for gesture trials $A$ and $B$, respectively.

Finally, we compute the consistency score following the definition proposed by Vatavu and Ungurean [VU22] using pairwise MDTW distances between gesture trials. For each gesture trial $g$, consistency was defined as:

$$\text{AC}(g) = \max\left(0,\ 1 - \frac{\text{average}_h\{\delta(g,h)\}}{\text{average}_{h^*}\{\delta(g,h^*)\}}\right) \tag{6.9}$$

where $h$ shows trials of the same gesture type (like g), $h^*$ shows trials of different gesture types, and $\delta$ represents the MDTW distance. This metric quantifies how similar each trial is to other instances of the same gesture type (intra-gesture similarity) versus how different it is from other gesture types (inter-gesture dissimilarity).

### 6.1.3 Gesture Recognition Baseline

We trained LSTM-based classifiers as baseline models to benchmark the suitability of MAGIC for gesture classification under both user-dependent and user-independent conditions. LSTMs are well-suited to multivariate time-series classification and have been widely applied to skeleton-based gesture and action recognition [? ? ]. We evaluated two input representations derived from the 3D motion-tracking data. The first baseline uses raw 3D sensor positions and velocities. In this representation, we use the palm sensor together with the fingertip and proximal finger sensors. For each trial, the recorded trajectories were first translated relative to the mean 3D position of the hand-rest reference sensor, to remove absolute positional differences across trials. Velocity was computed from the translated positional trajectories prior to resampling. Positional and velocity trajectories were then jointly resampled to a fixed length of 150 samples using arc-length resampling . Each input sequence therefore consists of 150 time steps, where each time step contains the combined 3D positions and velocities of the palm, fingertip, and proximal finger sensors. The second baseline uses a handcrafted feature representation combining hand-pose features and palm trajectory features described in Section 6.1.2. Each input sequence likewise consists of 150 time steps, where each time step contains the combined 15-dimensional hand-shape and 3-dimensional palm-trajectory features.

Both baselines share the same LSTM architecture. The network consists of one LSTM layer with 128 hidden units, followed by dropout with a rate of 0.3 and two fully connected layers with a ReLU activation between them, mapping the LSTM's 128-dimensional hidden state to the 18 gesture classes, followed by a softmax output. The model is trained using the Adam optimizer with a learning rate of $5 \times 10^{-4}$ and weight decay of $10^{-4}$, a batch size of 16, and cross-entropy loss, for a maximum of 120 epochs. Gradient clipping with a maximum norm of 1.0 is applied during training, and early stopping (patience of 10 epochs) is used based on validation macro-F1 score. We evaluated both representations under within-participant and between-participant conditions. For within-participant recognition, the trials of each participant were split into training, validation, and test subsets (20% test hold-out, with 10% of the remaining training data used for validation), and a single model was trained on the combined training data from all participants. For between-participant recognition, we used leave-one-participant-out (LOPO) evaluation, where each participant was held out for testing, three of the remaining participants were randomly selected for validation, and the rest formed the training set.

### 6.1.4 Gesture Consistency in Learned Deep Feature Space

We analyze gesture consistency in the learned embedding space of the LSTM classifier trained on the position-and-velocity representation. For each gesture trial, we extract the final hidden state of the LSTM layer as a 128-dimensional sequence-level embedding. Pairwise Euclidean distances are then computed between trial embeddings and used to calculate a consistency score following the same formulation as Section 6.1.2. For within-participant $\mathrm{AC}_{\mathrm{deep}}$, embeddings are extracted per participant, and pairwise distances are computed across that participant's own trials only. For between-participant $\mathrm{AC}_{\mathrm{deep}}$, the held-out participant's embeddings are compared against all trials from the remaining participants in the corresponding LOPO evaluation.

## 6.2 Validation of the Dataset and Evaluation of Gesture Execution Performance

We validated the MAGIC dataset by analyzing both intrapersonal and interpersonal variability in gesture execution, assessing consistency across hand-shape, palm trajectory, and learned deep feature representations, and establishing gesture recognition baselines using motion-related metrics described in Section 6.1. We first describe the statistical procedures used for data analysis. For within-subject comparisons across phases, we tested the normality of the paired differences using the Shapiro–Wilk test. When normality was confirmed, we used paired t-tests; otherwise, we applied Wilcoxon signed-rank tests. For between-subject comparisons, we analyzed differences in gesture performance across participants using Linear Mixed-Effects Models (LMMs) fitted on all trial-level data, with participant specified as a random effect. Additionally, to explore structure among participants, we applied Agglomerative Hierarchical Clustering to interpret subgroups based on characteristics of gesture execution. For gesture type comparisons, we performed one-way repeated-measures ANOVA or Friedman tests, depending on the data properties, followed by post hoc paired t-tests or Wilcoxon signed-rank tests with the Bonferroni correction.

### 6.2.1 Geometric Measures

**Quantity of Movement.** On average, gesture execution involved *1.34* meters of hand and arm movement (SD = 0.44). The amount of movement varied significantly by gesture type, $\chi^2(17) = 545.69$, $p < 0.001$. See Figure 6.1(a) for gesture-specific values. Post hoc comparisons (Wilcoxon signed-rank test with Bonferroni correction ($\alpha_{\text{adj}} = .005$)) revealed significant differences in QM among several gesture types. In particular, gestures such as *draw a rectangle*, *X symbol*, *wave hand*, and *grab hand and move forward* involved substantially larger movement quantities than many other gestures (all $p < .001$). In contrast, gestures such as *point right*, *lingering open hand*, *tap with index*, *OK sign*, and *thumbs up* were associated with significantly smaller movement quantities compared to some other gestures (all Bonferroni-corrected $p < .001$).

We also examined within-subject differences in gesture execution between the animation-guided phase (Phase 1) and the memory-based phase (Phase 2). Paired-sample t-tests revealed significant differences for several gestures. Specifically, quantities for *double tap with index* ($t(40) = -3.83, p < .001$), *point right* ($t(41) = -4.78, p = p < .001$), and *tap with index* ($t(41) = -2.95, p = p < .05$) were significantly smaller during Phase 1 compared to Phase 2. In contrast, for *wave hand* ($t(41) = 3.32, p < .001$), participants exhibited greater movement quantities during Phase 1 than during Phase 2.

**Gesture Movement Distribution.** To characterize the distribution of movement across the upper limb, we analyzed the relative contribution of the arm (Elbow, Humerus) and shoulder during gesture execution. On average, the proximal segments (arm and shoulder) accounted for 35% (SD = 18%) of total upper-limb movement, indicating that most gestures were executed primarily through the distal segments (i.e., hand and fingers). Due to the minimal contribution of proximal segments to total movement, gesture volume and area were computed using only the trajectories of hand and finger sensors, which better represent the space actively used during gesture execution.

**Gesture Area and Volume.** Gestures occupied an average frontal view area of *0.24 $m^2$* (SD = 0.1) and an average volume of *0.08 $m^3$* (SD = 0.5). See Figure 6.1(b,c) for the average area and volume of movement of the individual gestures. We found a significant effect of gesture type on both area ($\chi^2(17) = 528.99$, $p < 0.001$) and volume ($\chi^2(17) = 493.29$, $p < 0.001$).

Gestures such as the *X symbol*, *draw a rectangle*, *wave hand*, and *swipe hand* occupied significantly greater frontal area and volume compared to many other gestures. In contrast, gestures such as *point right*, *lingering open hand*, *OK sign*, and *double tap with index* occupied significantly smaller areas and volume. These differences were significant in most pairwise comparisons ($p < 0.005$, Bonferroni-corrected).

We examined within-subject differences in gesture execution between Phase 1 and Phase 2 of the study. For gesture area, significant differences were found for the following gestures: *double tap with index* ($W = 133$, $p < .001$), *tap with index* ($t(41) = -3.54$, $p < .001$), *show number 3* ($t(41) = -2.47$, $p < .05$), and *point right* ($t(41) = -5.3$, $p < .001$). In all cases, the gesture area in Phase 2 was significantly greater than in Phase 1. The same pattern held for volume, with significant increases in Phase 2 observed for *double tap with index*, *tap with index*, *show number 3*, *point right*.

The between-subject analysis for geometric features revealed higher participant effects compared to kinematic features. For quantity of movement, ICCs ranged from 0.52 to 0.83 (M = 0.69, SD = 0.09); for gesture area, from 0.60 to 0.82 (M = 0.72, SD = 0.07); and for gesture volume, from 0.50 to 0.83 (M = 0.71, SD = 0.09). These results indicate that gesture quantity and movement range were relatively consistent within individuals but varied more substantially between participants. Agglomerative hierarchical clustering revealed that Volume and QM were the most interpretable participant clusters, which aligned with demographic characteristics such as height and self-reported comfort with mid-air gestures. For instance, participants in the high-volume cluster had a mean height of 174 cm, compared to 163 cm in the low-volume cluster, and also reported higher gesture comfort ratings (mean 3.1 vs. 1.8).

### 6.2.2 Kinematic measures

**Performance Time and Speed.** Gestures took on average *2.95* seconds to execute (SD = 0.88), at an average speed of *47.14* cm/sec (SD = 14.01); See Figure 6.2 for the average production times and speeds of the individual gestures. We found a significant effect of gesture type on performance time ($\chi^2(17) = 322.58$, $p < 0.001$), as well as on users' average speed ($\chi^2(17) = 429.0$, $p < 0.001$).

Gestures such as *draw a rectangle*, *grab hand and move forward*, and *X symbol* exhibited significantly longer execution times compared to most other gestures. Speed also varied significantly across gesture types (reported above). However, because our speed measure (S) reflects gross upper-limb velocity rather than gesture-specific speed, we report it descriptively Figure 6.2(b) and do not compare it across gesture types.

Within-subject comparisons between Phase 1 and Phase 2 used paired t-tests or Wilcoxon signed-rank tests depending on normality. In terms of speed *double tap with index* (t(40) = -4.29, p < 0.001), *move apart thumb and index finger* (t(41) = -2.43, p < 0.05), *point right* (t(41) = -4.01, p < 0.001), *tap with index* (t(41) = -7.95, p < 0.001), and *turn knob* (t(41) = -3.47, p < 0.001) showed significant differences based on paired t-tests. *x symbol* (W = 221, p < 0.001) and *swipe hand* (W = 281, p < 0.05) showed significant differences according to Wilcoxon signed-rank tests. For all these gestures, speed was significantly higher in Phase 2 than in Phase 1.

In terms of performance time, significant differences between phases were observed for *Swipe hand* (t(41) = 2.91, p < 0.001), *tap with index* (t(41) = 7.38, p < 0.001), and *x symbol* (t(41) = 2.28, p < 0.05) showed significant differences based on paired t-tests. *Number 3* (W = 272, p < 0.05), *turn knob* (W = 102, p < 0.001), and *wave hand* (W = 192, p < 0.001) showed significant differences according to Wilcoxon signed-rank tests. For all these gestures, execution time was significantly shorter in Phase 2 than in Phase 1.

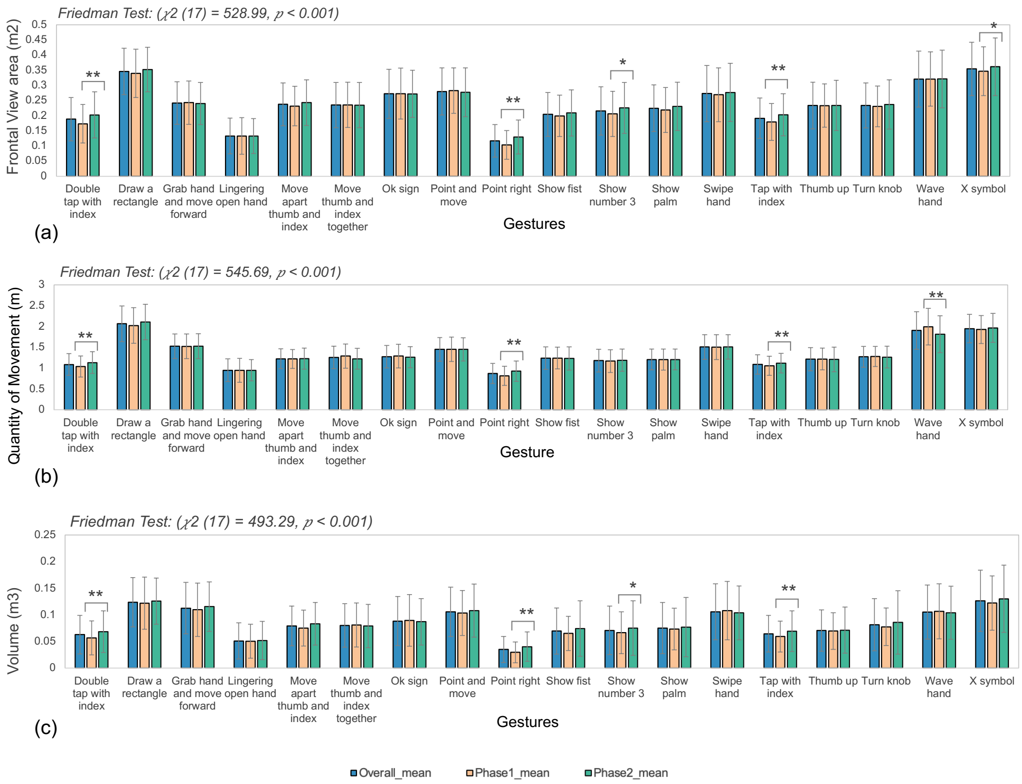



**Figure 6.1:** Mean geometric measures for each gesture, including a) area, b) quantity of movement, and c) volume, across Phase 1, Phase 2, and the overall condition (aggregated across phases). Error bars indicate standard deviation. Pairwise comparisons between Phase 1 and Phase 2 were conducted for each gesture; asterisks denote significant differences (* $p < 0.05$, ** $p < 0.001$).

The between-subject analysis of gesture performance time showed ICC values ranged from 0.14 to 0.64 (M = 0.47, SD = 0.12), suggesting that individual differences accounted for roughly 14–64% of the total variance. For gesture speed, ICC values ranged from 0.43 to 0.63 (M = 0.55, SD = 0.06), indicating that participant differences accounted for about half of the observed variability in movement speed across gestures. However, the clusters formed by time and speed did not correspond to any demographic characteristics, suggesting that temporal execution patterns may reflect individual movement styles rather than the demographic factors we considered.

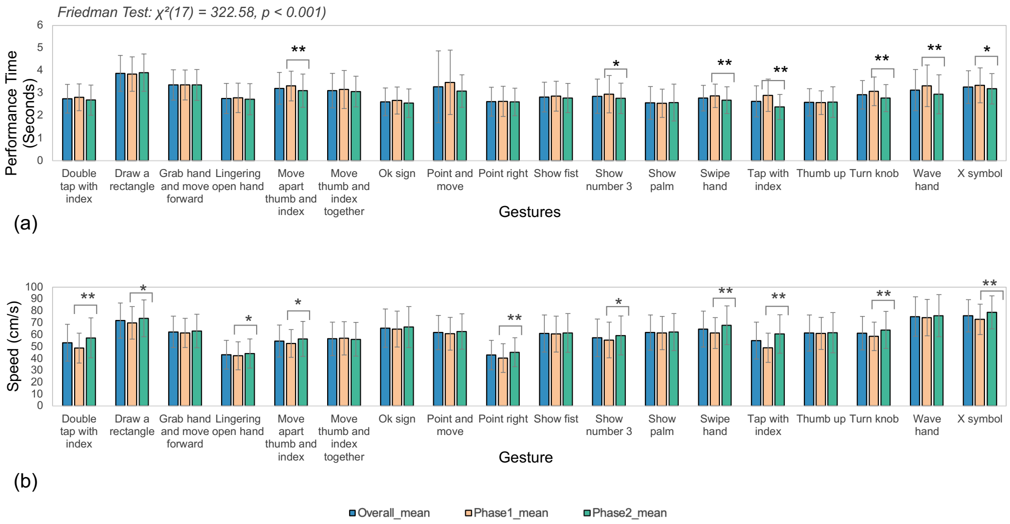



**Figure 6.2:** Mean kinematic measures for each gesture, including (a) performance time (PT) and (b) movement rate (S), across Phase 1, Phase 2, and the overall condition (aggregated across phases). Error bars indicate standard deviation. Pairwise comparisons between Phase 1 and Phase 2 were conducted for each gesture; asterisks denote significant differences (* $p < 0.05$, ** $p < 0.001$).

### 6.2.3 Gesture Performance Consistency

We assessed gesture performance consistency using hand shape ($AC_{shape}$), palm trajectory ($AC_{traj}$), and learned deep features ($AC_{deep}$), as described in Section 6.1.2. Table 6.1 reports within- and between-participant scores for all 18 gesture types across all three representations. We computed pairwise MDTW distances among all gesture recordings, generating a dissimilarity matrix that captures both within-gesture similarity and between-gesture dissimilarity, which forms the basis for the $AC_{shape}$ and $AC_{traj}$ scores reported below. Figure 6.3 presents an NMDS ordination of these dissimilarities based on hand-shape features, revealing four broad gesture groupings based on hand configuration: extended-index gestures, thumb and index coordination gestures, open-hand and grasp-like gestures, and two kinematically isolated gestures *thumb up* and *point right* which remained distinct across all dimensions. Hand shape was used for this visualization as it provides the most interpretable basis for characterizing gesture clusters by hand configuration.

or $AC_{shape}$, the average score across all gesture types was 0.50 (SD=0.15). *Thumb up*, *show palm*, *lingering open hand*, *wave hand*, and *show number 3* achieved the highest scores, while *grab hand and move forward*, *show fist*, *move thumb and index finger together*, and *move apart thumb and index finger* had the lowest.

At the participant level, within-participant $AC_{shape}$ (M=0.65, SD = 0.14) was higher than between-participant $AC_{shape}$ (M = 0.5, SD=0.12), with values mostly ranging between 0.5 and 0.7 with no extreme outliers, indicating that individuals perform gestures more similarly to their own repetitions than to those of other participants.

$AC_{traj}$ revealed a different pattern from $AC_{shape}$. Static or pose-dominant gestures such as *show palm*, *lingering open hand*, *point right*, and *thumb up* showed high $AC_{shape}$ ($\geq 0.68$) but low $AC_{traj}$ ($\approx 0.10$ within, $\leq 0.12$ between), as their palm displacement is minimal and does not reflect meaningful differences in execution. In contrast, gestures with a well-defined spatial path produced the highest $AC_{traj}$ scores: *draw a rectangle*, *x symbol*, *wave hand*, and *swipe hand*. This shows that for trajectory-defined gestures, consistency lies in their spatial path rather than their instantaneous hand configuration, a distinction that $AC_{shape}$ alone would miss. At the participant level, within-participant $AC_{traj}$ ($M = 0.27$, $SD = 0.15$) and between-participant $AC_{traj}$ ($M = 0.21$, $SD = 0.15$) were both substantially lower than the corresponding $AC_{shape}$ values, which reflects the greater variability in palm trajectory across repetitions compared to hand configuration.

For $AC_{deep}$, within-participant scores (M=0.67, SD=0.07) were higher than between-participant scores (M=0.51, SD=0.08) across all gesture types, following the same direction as $AC_{shape}$. Among gesture types, *thumb up*, *lingering open hand*, *show number 3*, and *point right* achieved the highest between-participant scores, while *move thumb and index finger together*, *point and move*, and *turn knob* were the lowest. For gestures with very low $AC_{traj}$ but high $AC_{shape}$, such as *lingering open hand* and *show palm* $AC_{deep}$ also remained high.

### 6.2.4 baseline gesture recognition Performance

The LSTM classifier trained on the position-and-velocity representation achieved a between-participant (LOPO) accuracy of 78.4% (mean F1 = 78.4%) and a within-participant accuracy of 83.7% (mean F1 = 83.0%) across all 18 gesture types. Per-participant LOPO accuracy varied across individuals, ranging from 60.9% (P30) to 95.4%, with the majority of participants falling between 70% and 85%. Among the five lowest-performing participants in LOPO (53.4 %–58.7 %), four showed high within-participant accuracy (77.8 % and 83.3 %) alongside their low LOPO scores. The remaining one showed similarly low accuracy (within = 72.2 %), reflecting fundamental execution variability rather than idiosyncratic style. Using the handcrafted feature representation, the between-participant accuracy was 72.6% (mean F1 = 72.2%) and the within-participant accuracy was 85.0% (mean F1 = 84.9%).

Figure 6.4 presents the LOPO confusion matrix and per-gesture F1-scores for both representations. Recognition performance varied substantially across gesture types. *Thumb up* (F1 = 0.92), *lingering open hand* (F1 = 0.90), *show palm* (F1 = 0.91), *wave hand* (F1 = 0.90), and *swipe hand* (F1 = 0.87) achieved the highest recognition rates, consistent with their stable hand configurations and high between-participant $AC_{shape}$ scores (Section **??**). The lowest recognition score were observed for *double tap with index* (F1 = 0.39), *tap with index* (F1 = 0.47), *move thumb and index finger together* (F1 = 0.56), and *move apart thumb and index finger* (F1 = 0.71). The most frequent confusion occurred between *double tap with index* and *tap with index*, accounting for 18.7% of all

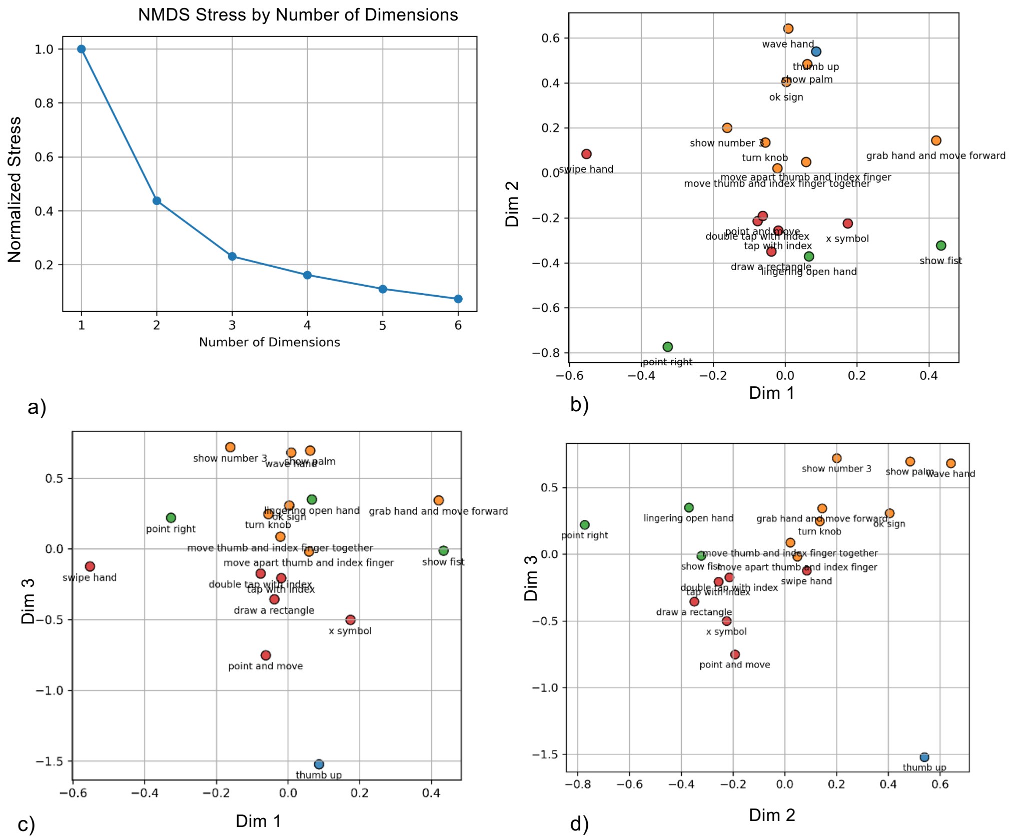



**Figure 6.3:** NMDS analysis of gesture articulation dissimilarities. a) Stress values by number of dimensions, b-d) NMDS ordinations (Dim 1 vs Dim 2, Dim 1 vs Dim 3, and Dim 2 vs Dim 3) based on MDTW dissimilarities of normalized hand-shape features.

**Table 6.1:** Within- and between-participant articulation consistency (AC) scores for each gesture type, computed using shape-based, trajectory-based, and learned deep representations.

| Gesture | AC (Shape, Within) | AC (Shape, Between) | AC (Trajectory, Within) | AC (Trajectory, Between) | AC (Deep, Within) | AC (Deep, Between) |
|---|---|---|---|---|---|---|
| **Double tap with index** | 0.42 | 0.61 | 0.14 | 0.09 | 0.56 | 0.46 |
| **Draw a rectangle** | 0.45 | 0.55 | 0.65 | 0.63 | 0.70 | 0.52 |
| **Grab hand and move forward** | 0.24 | 0.39 | 0.38 | 0.38 | 0.68 | 0.48 |
| **Lingering open hand** | 0.69 | 0.83 | 0.1 | 0.08 | 0.76 | 0.69 |
| **Move apart thumb and index** | 0.33 | 0.49 | 0.15 | 0.11 | 0.62 | 0.43 |
| **Move thumb and index together** | 0.31 | 0.47 | 0.16 | 0.14 | 0.55 | 0.38 |
| **Ok sign** | 0.59 | 0.64 | 0.16 | 0.11 | 0.66 | 0.50 |
| **Point and move** | 0.49 | 0.70 | 0.39 | 0.33 | 0.55 | 0.41 |
| **Point right** | 0.48 | 0.59 | 0.1 | 0.03 | 0.75 | 0.58 |
| **Show fist** | 0.29 | 0.57 | 0.14 | 0.09 | 0.69 | 0.48 |
| **Show number 3** | 0.60 | 0.74 | 0.15 | 0.06 | 0.76 | 0.54 |
| **Show palm** | 0.73 | 0.91 | 0.18 | 0.05 | 0.73 | 0.62 |
| **Swipe hand** | 0.42 | 0.71 | 0.4 | 0.42 | 0.64 | 0.54 |
| **Tap with index** | 0.45 | 0.64 | 0.2 | 0.09 | 0.61 | 0.49 |
| **Thumb up** | 0.72 | 0.78 | 0.21 | 0.12 | 0.76 | 0.65 |
| **Turn knob** | 0.48 | 0.67 | 0.2 | 0.11 | 0.65 | 0.42 |
| **Wave hand** | 0.68 | 0.81 | 0.48 | 0.41 | 0.69 | 0.53 |
| **X symbol** | 0.44 | 0.53 | 0.61 | 0.6 | 0.70 | 0.51 |

misclassifications, with 37.7% of *double tap with index* trials misclassified as *tap with index* and 33.9% of *tap with index* trials misclassified in the reverse direction. Excluding this pair, overall accuracy rises to 85.4% and mean F1 to 83.0%. Further confusion occurred among *move thumb and index finger together*, *move apart thumb and index finger*, and *ok sign*, with mutual error rates of 6.1 to 12.2% per pair, reflecting their shared thumb and index coordination dynamics. *X symbol* and *draw a rectangle* showed mutual confusion rates of 10.3% and 7.4%, and *grab hand and move forward* and *show fist* showed mutual confusion rates of 8.5% and 7.6%. These confusion patterns are consistent with the kinematic groupings observed in the NMDS ordination of gesture dissimilarities (Figure 6.3), where gestures that are frequently confused share close proximity in the ordination space.

Furthermore, to examine the relationship between gesture execution characteristics and recognition performance, we computed Pearson correlation coefficients between per-gesture LOPO F1 scores and each of the three AC representations as well as the five geometric and kinematic measures across the 18 gesture types. Among the AC representations, $AC_{deep}$ showed the highest correlation with F1 ($r(18) = .547$, $p = .019$). $AC_{shape}$ also showed a positive association with LOPO F1 scores ($r(18) = .407$, $p = .090$), suggesting that handcrafted measures partially capture cross-user recognition difficulty. $AC_{traj}$ did not correlate significantly with F1 ($r(18) = .289$, $p = .245$), consistent with the fact that trajectory consistency is only meaningful for a subset of gesture types. Among the geometric and kinematic measures, none reached statistical significance: quantity of movement ($r(18) = .313$, $p = .206$), gesture area ($r(18) = .256$, $p = .305$), gesture volume ($r(18) = .254$, $p = .309$), speed ($r(18) = .299$, $p = .229$), and performance time ($r(18) = .105$, $p = .680$), indicating that the physical demand of a gesture does not predict its cross-user recognizability in our normalized recognition pipeline.

### 6.2.5 Confidence Ratings and Memory Performance

In this section, we analyze participants' confidence ratings and memory recall performance of execution gestures. In the first phase, participants rated their gesture accuracy and confidence on a 10-point Likert scale. The results across gestures were highly consistent, with average accuracy ratings ranging from 8.3 to 9.5 (SDs < 1.1 for most gestures), and confidence ratings ranging from 7.9 to 9.1. The gesture *Show palm* received the highest accuracy (mean= 9.5, SD=0.65) score and confident score (mean = 9.1, SD = 0.75), while *move thumb index finger together* had the lowest confident score (mean = 7.9, SD = 1.11) and the *draw rectangle* had the lowest accuracy (mean = 8.3, SD = 1.92).

In the second phase of the study, the average confidence ratings ranged from 7.7 to 9.3 (SDs < 1.4 for most gestures), while memory ratings ranging from 6.5 to 9.4. The *Grab Hand and Move Forward* gesture received the highest memory score (mean = 9.4, SD = 2.10) and confidence (mean = 9.3, SD= 2.3), while *Lingering Open Hand* and *turn knob* had the lowest memory recall score (mean = 6.5, SD = 2.3) and (6.7 SD= 1.7).

Several participants also reported confusion between the *thumb up* and *OK sign* gestures, with approximately 9% failing to accurately recall and distinguish these two gestures on their first attempt. The *Lingering Open Hand* gesture was also frequently cited as difficult to remember, and participants substituted it with other open-hand gestures. Overall, about 12% of participants were unable to recall the *Lingering Open Hand* gesture. Additionally, some participants expressed uncertainty about the direction of movement of gestures involving motion trajectories, such as *Swipe Hand* and *Point and Move*, which was reflected in their confidence rating scores.

| | Double tap with index | Draw a rectangle | Grab hand and move forward | Lingering open hand | Move apart thumb and index finger | Move thumb and index finger together | Ok sign | Point and move | Point right | Show fist | Show number 3 | Show palm | Swipe hand | Tap with index | Thumb up | Turn knob | Wave hand | X symbol | F1 (position+velocity) | F1 (hand-crafted features) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Double tap with index | 90 | 0 | 0 | 8 | 6 | 30 | 0 | 7 | 10 | 1 | 0 | 0 | 0 | 87 | 0 | 0 | 0 | 0 | 0.39 | 0.35 |
| Draw a rectangle | 2 | 189 | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 16 | 0.87 | 0.79 |
| Grab hand and move forward | 0 | 0 | 189 | 4 | 2 | 2 | 0 | 0 | 0 | 20 | 6 | 0 | 1 | 0 | 1 | 9 | 2 | 0 | 0.80 | 0.77 |
| Lingering open hand | 4 | 0 | 2 | 222 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 5 | 0 | 0 | 0.90 | 0.93 |
| Move apart thumb and index | 10 | 1 | 0 | 3 | 161 | 27 | 20 | 0 | 0 | 0 | 0 | 2 | 0 | 11 | 0 | 0 | 0 | 0 | 0.71 | 0.45 |
| Move thumb and index together | 17 | 2 | 2 | 1 | 28 | 131 | 19 | 0 | 1 | 7 | 4 | 0 | 0 | 14 | 0 | 1 | 0 | 2 | 0.56 | 0.35 |
| Ok sign | 1 | 0 | 1 | 0 | 10 | 14 | 184 | 1 | 0 | 2 | 0 | 6 | 0 | 0 | 1 | 0 | 2 | 2 | 0.80 | 0.81 |
| Point and move | 9 | 2 | 2 | 0 | 0 | 2 | 0 | 187 | 8 | 1 | 0 | 0 | 7 | 6 | 0 | 0 | 0 | 4 | 0.84 | 0.74 |
| Point right | 5 | 2 | 2 | 2 | 0 | 0 | 0 | 8 | 208 | 0 | 0 | 0 | 2 | 9 | 0 | 3 | 0 | 1 | 0.86 | 0.86 |
| Show fist | 3 | 0 | 18 | 2 | 0 | 3 | 0 | 0 | 1 | 195 | 2 | 0 | 0 | 2 | 8 | 3 | 0 | 0 | 0.81 | 0.76 |
| Show number 3 | 0 | 0 | 7 | 3 | 0 | 1 | 9 | 0 | 0 | 0 | 202 | 13 | 0 | 0 | 0 | 4 | 2 | 0 | 0.85 | 0.87 |
| Show palm | 0 | 0 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 219 | 0 | 0 | 0 | 4 | 1 | 0 | 0.91 | 0.90 |
| Swipe hand | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 206 | 0 | 2 | 7 | 20 | 1 | 0.87 | 0.85 |
| Tap with index | 80 | 0 | 1 | 3 | 5 | 24 | 0 | 2 | 7 | 2 | 0 | 0 | 0 | 112 | 0 | 0 | 0 | 0 | 0.47 | 0.36 |
| Thumb up | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 0 | 0 | 3 | 0 | 218 | 3 | 3 | 0 | 0.92 | 0.95 |
| Turn knob | 6 | 2 | 9 | 2 | 0 | 2 | 0 | 0 | 8 | 4 | 7 | 2 | 2 | 2 | 1 | 182 | 2 | 0 | 0.80 | 0.71 |
| Wave hand | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 9 | 0 | 0 | 6 | 225 | 0 | 0.90 | 0.84 |
| X symbol | 1 | 23 | 0 | 0 | 1 | 1 | 0 | 5 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 190 | 0.86 | 0.71 |

(a) Low High (b)

**Figure 6.4:** LOPO Gesture recognition baseline results. (a) Confusion matrix for the position-and-velocity representation. (Right) Per-gesture F1-scores for the position-and-velocity and handcrafted feature representations.

## 6.3 Discussion

We analyzed our dataset containing 18 mid-air gestures, which was described in the previous chapter. Our analysis of the dataset reveals variability in gesture consistency, both across users and between gesture types, as well as differences in gesture kinematics and geometry performance. We also observed how users adapted their gestures over time, with practice leading to increased speed without compromising consistency for the majority of gestures. Also, certain gesture categories required greater physical effort or exhibited lower articulation consistency than others. Together, these findings offer new insight into the dynamics of mid-air gesture execution and provide an empirical basis for evaluating gesture suitability in different interactive systems. In the remainder of this section, we discuss these findings in detail and consider their implications for interface design, recognition models, and future work.

### 6.3.1 Gesture Performance and Variability

Our analysis shows that gesture performance varies across users, gesture types, and, for some gestures, study phases. $AC_{shape}$ was highest when all fingers reached a well-defined, anatomically constrained endpoint, as in static gestures like *show palm* and *thumb up* or dynamic ones like *swipe hand*. It was lowest for gestures involving partially specified finger configurations like *grab hand and move forward*, where the degree of closure is not semantically defined, and *move thumb and index together or away*, where three non-critical fingers remain free. Notably, the latter two achieve moderate LOPO recognition rates despite their low consistency scores, suggesting that variability in non-critical fingers does not substantially impair recognition when the semantically active fingers move consistently. $AC_{traj}$ acquired meaningfully above zero only for gestures defined by a spatial path, such as *draw a rectangle*, which, despite low $AC_{shape}$, achieved high LOPO recognition rates, indicating that trajectory consistency can sustain reliable cross-user recognition even when hand configuration varies.

### 6.3.2 Kinematic and Spatial Demands in Gesture Design

Gesture geometry has important implications for user experience, particularly in frequent or prolonged interaction. In our dataset, gestures varied significantly in execution time, movement speed, and spatial footprint. Trajectory-based gestures, such as *draw a rectangle*, took longer to perform and occupied substantially larger spatial areas and volumes than static pose-based gestures such as *thumb up*, which were more spatially compact. Trajectory-based gestures may not be suitable for high-frequency commands, but can be effective for occasional and expressive actions in which visibility and clarity are prioritized. The thumb-up gesture, on the other hand, was both fast and spatially compact, which under our proxy measures, suggests a lower physical cost and a greater suitability for frequent interactions. For systems requiring repeated gesture input, such as gesture-controlled kiosks or smart home interfaces, gesture vocabularies should reserve gestures with higher proxy-based physical demand such as *draw a rectangle* for low-frequency expressive actions, while assigning frequent commands to compact, low-demand gestures such as *thumb up* or *tap with index*. We did not collect direct measures of exertion, discomfort, or muscle activity; validating these proxy-based estimates against such measures (e.g., Borg-scale ratings or EMG) is an important direction for future work. A key distinction arises between gestures involving large arm movements and those requiring fine motor control. Gestures such as *move thumb and index finger together* showed among the lowest QM, which might suggest they are less physically demanding than trajectory-based gestures. However, because QM aggregates movement across all upper-limb sensors and is dominated by proximal segment displacement, these measures reflect gross upper-limb demand rather than overall gesture effort. For gestures whose physical demand is concentrated in the fingers rather than the arm, low kinematic values are therefore interpreted as indicating small upper-limb displacement rather than low effort.

### 6.3.3 Public vs. Private Systems

Our analysis revealed that between-person consistency was lower than within-person consistency, and this difference was reflected in recognition performance within our 18-gesture vocabulary. Some gestures, such as *thumb up* or *show palm*, achieved the highest LOPO recognition rates, while some gestures, such as *double tap*, showed lower recognition performance within this gesture vocabulary. For the public, walk-up systems, where recognition must generalize across diverse users, vocabulary selection is critical. Including gestures that are easily confused with others in the set increases the likelihood of triggering the wrong command. In contrast, private systems offer the opportunity to expand the gesture vocabulary beyond public deployment. Even gestures poorly suited for public vocabularies, such as *grab hand and move forward*, achieved within-participant accuracy above 77% for the majority of participants, making them viable candidates for personalized systems.

The confusion analysis further highlights that gesture pairs differing only in temporal repetition, such as *tap with index* and *double tap with index*, are particularly problematic for user-independent systems, suggesting that gesture vocabularies should avoid gestures distinguished solely by timing when cross-user robustness is required.

Moreover, our correlation analysis showed that $AC_{shape}$ was positively associated with LOPO F1 scores (r=.407r = .407 r=.407), suggesting it can serve as a lightweight indicator for vocabulary design without requiring a full recognition pipeline. In contrast, geometric and kinematic measures did not predict recognition performance.

### 6.3.4 Scaling Gesture to Human Variability

Gestures exhibited significant variation in execution time, speed, and movement extent across participants. This variation was driven by different factors for geometric and kinematic features. Spatial characteristics such as movement volume and extent reflected physical body dimensions and individual disposition toward gesture use (i.e., taller participants and those who reported higher comfort with gestural interaction tended to perform gestures more expansively). In contrast, temporal characteristics such as execution speed and timing did not correspond to any demographic or anthropometric variable, suggesting they reflect individual motor styles rather than physical body dimensions or fixed group traits such as gender, as no distinct participant groups were observed .

This has design implications for inclusive systems that need to scale across populations. Rather than relying on fixed geometric or temporal cut-offs that may disproportionately penalize, for example, smaller-bodied users or those with limited mobility [YPFK+23; CTCG17], especially for trajectory-rich gestures, interaction designs should account for population diversity through established strategies such as design-for-extreme, adjustable design, design for average, and design-for-robustness approaches that seek to maintain usability and performance across different users [WGLL04]. Our dataset provides contributions to support such strategies by quantifying the distribution and variability of gesture execution parameters.

From an accessibility perspective, this also raises concerns around fatigue. Gestures like *draw a rectangle* or *grab and move forward* carry higher kinematic demands under our proxy measures, which may lead to quicker physical exhaustion if used repetitively. Recognition systems should therefore not only measure performance accuracy, but should also take into account how physically demanding each gesture appears based on these proxy measures, possibly adjusting the system to avoid tiring gestures or lowering the confidence threshold when a gesture is harder to perform repeatedly.

### 6.3.5 Leveraging Lightweight Training for Efficient Gesture Input

We found that after brief practice and removal of visual guidance, participants performed several gestures faster, with shorter execution times and, in some cases, larger spatial footprints. Importantly, consistency remained stable across phases, indicating that gains in fluency did not compromise structural reliability. Gesture-based interfaces can benefit from this by providing lightweight onboarding, such as interactive tutorials, to facilitate rapid user familiarization and promote consistent, reliable input. Interfaces can anticipate that users will execute gestures more quickly and sometimes with larger movement envelopes after initial exposure. Systems should therefore allow adaptive thresholds for timing and spatial extent to accommodate this motor adaptation without increasing misclassification risk, and should also consider potential fatigue costs when gestures expand in size, especially for commands expected to be performed frequently.

## 6.4 Limitations and Future Work

Our geometric and kinematic measures are indirect proxies for physical demand and were not validated against direct ergonomic measures such as perceived exertion (e.g., Borg CR10), discomfort ratings, or muscle activity (EMG). Future work should address these limitation and incorporate direct ergonomic validation of the proxy measures used here.

### Conclusion

We presented MAGIC, an open-access dataset of 18 mid-air gestures collected from 42 participants, with multiple repetitions per gesture, recorded synchronously via RGB, depth, and high-precision motion tracking. We validated it through analyses of within-person and between-person variability, kinematic/geometric demands, and articulation consistency using MDTW-based metrics. We found that certain static pose-based gestures, such as thumb up and show palm, achieved higher articulation consistency and lower spatial demand, while trajectory-rich gestures, such as draw a rectangle and grab hand and move forward, required more time and space and showed greater variability across users. Despite this, participants were more consistent in their own repetitions, and short-term practice improves fluency (faster, sometimes larger trajectories) without reducing consistency. With MAGIC, we provide the vision-based and motion-tracking research communities with a common benchmark dataset for assessing the applicability and robustness of gesture execution and recognition approaches.

# 7 General Conclusions

In the final chapter, we conclude this thesis. First, we provide a summary of the work presented, which aimed to advance the understanding and design of mid-air gesture interactions in human–computer interaction. We revisit the research questions posed in Chapter 1 and highlight the key contributions made across gesture elicitation, vocabulary consistency, and the creation of a multimodal dataset for gesture performance analysis. Finally, we discuss the limitations of our approach and outline directions for future work in gesture-based interaction design, motion analysis, and dataset development.

## 7.1 Major Contributions to RQ1

With RQ1, we looked towards the consistency of user-defined mid-air gestures across different application domains in HCI, and how this informs the development of a transferable and consensus gesture set. Gesture interaction in HCI has largely developed within isolated, domain-specific studies, leading to fragmented vocabularies that limit transferability and reuse.

**RQ1: To what extent are mid-air gesture sets consistent across application domains, and how can this inform the design of a consensus gesture set transferable across domains?**

Our exploration began with a systematic literature review of 172 papers on mid-air gestural interaction (Chapter 3). Our goal was to determine if gesture vocabularies developed in specific contexts, such as entertainment, virtual reality, robotics, or smart environments, exhibit signs of transferability and consistency across domains. From the reviewed literature, we extracted gesture sets reported across studies and consolidated them into a unified dataset for cross-domain analysis. Each gesture was coded along multiple taxonomic dimensions, allowing for clustering based on similarities in both form and meaning. Consistency across domains was then assessed by applying an agreement-rate analysis, which measured the extent to which gestures that appeared for the same command were similar. From this analysis, 22 gestures were identified as highly consistent across domains and were proposed as candidates with strong potential for cross-domain transferability. Manipulative gestures, such as rotation or scaling, demonstrated the highest levels of consistency across domains due to the direct spatial correspondence between movement and effects. Symbolic gestures, such as swipe or circle, appeared frequently but exhibited high semantic variability, often representing different commands across domains (e.g., next, delete, or scroll). Overall, gesture consistency was strongly linked to the type of control being performed. It was higher when gestures targeted spatial objects or continuous control tasks, and lower when associated with abstract system commands.

In Chapter 4, we extended this investigation with a scenario-based elicitation study in a fully equipped smart home lab. Here, participants proposed gestures for real-world scenarios such as cooking, entertainment, fitness, and home office work. The analysis showed that users often reused and adapted familiar gestures such as swipe, rotate, push/pull, and zoom, retained high agreement across devices and scenarios when transitioning between scenarios. The results confirmed that gestures with strong spatial and semantic mappings retained high agreement across devices and contexts.

Taken together, these studies answer RQ1 by demonstrating that user-defined mid-air gestures are not isolated within individual applications but show consistency across domains. They provide both empirical evidence and methodological foundations for designing a cross-domain, transferable gesture vocabulary in HCI that allows users to build on existing gestural knowledge rather than relearn gestures for each application.

## 7.2 Major Contributions to RQ2

With RQ2, we examined the preferences and motivations that drive users to employ gesture-based interactions in smart home environments, and how these factors differ across scenarios. Unlike RQ1, which focused on cross-domain gesture consistency, RQ2 investigated the user perspective in terms of why and where gestures are perceived as useful or desirable, and how contextual demands shape their use.

**RQ2: How do users' preferences and motivations for gesture-based interactions in smart homes differ dependent to the scenarios?**

In Chapter 4, we addressed RQ2 through a framed guessability study conducted in four distinct scenarios within a smart home environment. The study immersed participants in four everyday contexts, including cooking, entertainment, home office, and fitness training. Each scenario was chosen for its contrasting characteristics, including different physical postures (e.g., standing, sitting, physical movement) as well as diverse activities and roles undertaken by users. Within each scenario, participants reflected on their routines, discussed motivations for using gestures, and proposed interactions for devices and commands. This allowed us to capture both their preferences (i.e., which devices and tasks they wished to control through gestures) and their underlying motivations (i.e., why they would or would not use gestures in that context).

From this analysis, the kitchen was perceived as the most promising environment for mid-air gesture interaction, followed by entertainment, while office and fitness contexts were considered less critical. In addition, a limited number of gesture vocabulary (approximately 5–9 gestures) was identified as optimal for preventing cognitive overload across the smart home ecosystem.

The results also showed that motivations for adopting mid-air gestures are scenario-dependent. Rather than expressing a universal preference for gestural interaction, gestures were valued they addressed specific barriers dependen to the scenario (e.g., avoiding device contact with dirty or wet hands while cooking), minimize disruption (e.g., during exercise), or enhance comfort and enjoyment (e.g., maintaining continuity during exercise), or enhance comfort and enjoyment (e.g., controlling entertainment systems from a relaxed position).

In addition, Gesture interaction was most suitable for frequently used household devices, such as lighting, media systems, and blinds, where gesture interactions primarily emerged in simple, high-frequency tasks performed at a distance (e.g., turning on or off, opening or closing, or adjusting volume and brightness). In contrast, tasks that required fine-grained manipulation or multi-step input, or precise control, were rarely mapped to gestures. In these cases, alternative input modalities such as touch inputs, voice assistants, or remote controls, were considered more efficient and reliable.

## 7.3 Major Contributions to RQ3

**RQ3: How are mid-air gestures physically performed across users and repetitions, and what do these performance characteristics reveal about their suitability for designing robust gesture-based interaction systems?**

With RQ3, we sought to go beyond the descriptive identification of gesture vocabularies to investigate how users actually perform gestures in practice. Prior elicitation and interaction design studies have revealed which gestures users find intuitive or appropriate, but they typically capture only intended gestures (e.g., sketches, descriptions, or video clips) rather than executed gesture trajectories. This leaves open questions about the physical feasibility, kinematic consistency, and ergonomic suitability of gestures across diverse users, as well as their implications for the reliable detection and recognition of systems.

To address these gaps, we developed the MAGIC dataset, a high-fidelity, open-access corpus capturing 18 mid-air gestures performed by 42 participants across multiple repetitions in two phases: a guided phase with visual instruction, followed by an unguided phase in which gestures were recalled and performed from memory, using synchronized RGB, depth, and electromagnetic motion-tracking modalities to capture both visual and kinematic data. In Chapter 5, we presented the development of the MAGIC dataset. Unlike existing datasets that are designed for specific tasks, such as dance, sports, or sign language, MAGIC provides a capture of the gestures most frequently proposed in HCI research across domains.

This dataset formed the foundation for the subsequent analysis of gesture performance characteristics presented in Chapter 6, where we quantified gesture performance using geometric measures (movement extent, frontal area, and spatial volume) and kinematic measures (execution time and movement speed). Variability in gesture articulation was then analyzed using Multidimensional Dynamic Time Warping to assess temporal and spatial consistency both within and across users.

According to the analysis of articulation, some gestures formed well-separated clusters with high inter-user consistency, making them reliable primitives for public walk-up interfaces with low false-positive tolerance. However, other gestures demonstrated lower inter-user consistency, for example, tap and double-tap. Despite this, since intra-user articulation consistency remeind high, i.e., individuals tend to reproduce their own gestures reliably across repetitions, these gestures can still effectively be used in personalized or household-specific models that support a broader vocabulary without compromising reliability. Furthermore, Short-term practice, from guided to memory-based execution, led to faster and reduced execution time without degrading articulation consistency, suggesting that lightweight onboarding and adaptive recognition thresholds can enhance both usability and system robustness.

## 7.4 future work

The work presented in this dissertation is subject to certain limitations, which can be addressed and improved in future work. While individual chapters addressed study-specific constraints, I synthesize the overall limitations and propose directions for extending this line of research in this section.

### Datasets and Participant Limitations

Across this thesis, datasets and participant groups played a central role in shaping the results. In Chapter 3, the corpus of surveyed publications was limited by the choice of query method, inclusion criteria, and the subscription databases available at the time of search in April 2021, which may have narrowed the scope by excluding short papers and later publications. In Chapter 4, participants were recruited from Central Europe, which limited cultural diversity in gesture elicitation and interpretation. Finally, in Chapter 5 and Chapter 6, the developed dataset involved participants performing gestures with their right hand and without representation of specific user groups, such as elderly users or individuals with motor impairments. These may limit the extent to which the results can be applied. To build datasets that better reflect diverse populations, future research could involve participants from different age groups, cultural backgrounds, and physical abilities. Similarly, future surveys could integrate additional venues, such as open-access databases and workshops, to provide a broader scope.

### Lack of Empirical Validation

Across this thesis, several contributions focused on eliciting gestures, identifying consensus sets, and developing datasets; however, these remain to some extend unvalidated in real-world scenarios. In Chapter 4, while participants proposed gestures and preferences for mid-air interactions in a smart home context, these gestures were not empirically tested for usability, efficiency, or long-term adoption. In Chapter 3, the consensus set of gestures derived from prior elicitation studies has yet to be examined under everyday conditions, where social, spatial, and cognitive factors may influence performance. Similarly, in Chapter 5, while a multi-modal dataset of gestures was developed, it has not been validated through real-world deployments or across varied physical and social environments. These limitations restrict the extent to which the proposed gestures and datasets can be confidently generalized to practical applications. Future work should therefore conduct empirical validation in naturalistic environments, assessing not only recognition accuracy but also user satisfaction, social acceptability, and long-term integration into daily activities.

### Integration with Multimodal Interaction

While this thesis focused on establishing gesture consistency and transferability across domains, future research should extend these findings toward multimodal interaction in pervasive environments. The input modalities are highly complementary and can mitigate the limited expressivity of single-modality systems, while enhancing speed, accuracy, and interaction flexibility across contexts [CXH15]. Building on the gesture sets defined in this work, future studies could explore how multimodal systems enable more natural, context-aware, and scalable interactions across diverse computing environments.

# Appendices

# Gesture Frequencies Across Domains

**Table A.1:** List of 3 most frequent gestures for each referent and their distribution between different domains. frequency $< 25\%$ ○. $25\% \leq$ frequency $< 50\%$ ⊙. $50\% \leq$ frequency $< 75\%$ ◉. frequency $\geq 75\%$ ●.

| | | Cross-Domain | 3D modeling | Accessibility and Adaptive System | eLearning and Education | Gaming | HRI | In-Vehicle and Automotive | Large (public) Display | Manipulation/ navigation | Media and Entertainment | Medical Technology | Mobile interaction | Smart Rooms / Appliances |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Translate | Swipe/Slide (63.7% of 135 gestures in 59 papers) | ◉ | ◉ | | ● | ◉ | ● | ⊙ | ◉ | ◉ | ◉ | ⊙ | ● | ● |
| | Grab and Move (20.74% of 135 gestures in 59 papers) | ○ | ○ | ● | | ⊙ | | ⊙ | ⊙ | ⊙ | ⊙ | ⊙ | | |
| | Point (5.92% of 135 gestures in 59 papers) | ○ | ○ | | ○ | | | ◉ | ◉ | | | | | |
| Rotate | Turning movement (71.9 % of 121 gestures in 56 papers) | ◉ | ● | | ● | ● | ⊙ | ● | | ◉ | ● | ● | ◉ | |
| | Push and Pull (12.4 % of 121 gestures in 56 papers) | ○ | ○ | | | | ◉ | | | ○ | | | | |
| | Swipe/Slide (10.74 % of 121 gestures in 56 papers) | ○ | | | ○ | | | | ⊙ | ○ | | | ⊙ | |
| Next | Swipe/Slide (71.83% of 71 gestures in 43 papers) | ◉ | | ⊙ | ◉ | | | ◉ | ◉ | ● | ◉ | | ◉ | ● |
| | Rotate (5.63% of 71 gestures in 43 papers) | ○ | | ⊙ | | | | | | | ○ | | ○ | |
| | Tap (5.63% of 71 gestures in 43 papers) | ○ | | | ◉ | | | | | | | | ⊙ | |
| Increase | Swipe/Slide ( 54.79% of 73 gestures in 42 papers) | ◉ | | ● | | | ● | ◉ | | | ◉ | | ● | ◉ |
| | Turn Knob/Rotate ( 16.44% of 73 gestures in 42 papers) | ○ | | | | | | ○ | | ◉ | ○ | | | ○ |
| | Move the Slider ( 10.96% of 73 gestures in 42 papers) | ○ | | | | | | ○ | | | ○ | | | ○ |
| Decrease | Swipe/Slide ( 62.85% of 70 gestures in 41 papers) | ◉ | | ● | | | ● | ◉ | | | ◉ | | ● | ● |
| | Turn Knob/Rotate ( 14.28% of 70 gestures in 41 papers) | ○ | | | | | | ○ | | ● | ○ | | | ○ |
| | Move/grab the Slider ( 10.00% of 70 gestures in 41 papers) | ○ | | | | | | ○ | | | ○ | | | ○ |
| Previous | Swipe/Slide ( 69.86% of 73 gestures in 41 papers) | ◉ | | ⊙ | | | | ◉ | ◉ | | ◉ | | ◉ | ● |
| | Rotate ( 7.04% of 73 gestures in 41 papers) | ○ | | ⊙ | | | | | ○ | | ○ | | ○ | |
| | Tap ( 4.22% of 73 gestures in 41 papers) | ○ | | | | | | | | | | | ⊙ | |
| Zoom | Pull/Squish (67.82% of 87 gestures in 39 paperss) | ◉ | ● | | ● | | ● | ● | ◉ | ● | ◉ | ◉ | ● | |
| | Swipe/Slide (12.64% of 87 gestures in 39 paperss) | ○ | | | | | | | | | ○ | | | |

*(continued on next page)*

*(continued from previous page)*

| | | Cross-Domain | 3D modeling | Accessibility and Adaptive System | eLearning and Education | Gaming | HRI | In-Vehicle and Automotive | Large (public) Display | Manipulation/ navigation | Media and Entertainment | Medical Technology | Mobile interaction | Smart Rooms / Appliances |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Move Closer/Move Away (12.64% of 87 gestures in 39 paperss) | ○ | | | | | | | ○ | ○ | ○ | ⊙ | | |
| Scale | Pull/Squish (88.31% of 77 gestures in 27 papers) | ● | ● | | ● | | | ● | ◉ | ● | ● | ● | | |
| | Swipe/Slide (3.89% of 77 gestures in 27 papers) | ○ | ○ | | | | | | | | | | | |
| | Move Closer/Move Away (3.89% of 77 gestures in 27 papers) | ○ | | | | | | | ◉ | ○ | | | | |
| Pan/scroll | Swipe/Slide (69.35% of 62 gestures in 31 papers) | ◉ | | | ● | | | ⊙ | ◉ | ⊙ | ● | | ● | |
| | Grab and Move (14.5% of 62 gestures in 31 papers) | ○ | | | | | | | ○ | ● | ⊙ | ● | | |
| | Rotate (6.45% of 62 gestures in 31 papers) | ○ | | | | | | ○ | | | | | | |
| Deselect | Open Hand (61.9% of 21 gestures in 16 papers) | ◉ | | ● | ● | ● | ● | | | ● | ○ | ● | | ● |
| | Point (4.5% of 21 gestures in 16 papers) | ○ | ● | | | | | | | | ○ | | | |
| | Thumb Down (4.5% of 21 gestures in 16 papers) | ○ | | | | | | | | | | | | |
| Close menu | Swipe/Slide (66.67% of 12 gestures in 8 papers) | ◉ | | | | | | ● | | ◉ | ● | | | |
| | Squish (8.33% of 12 gestures in 8 papers) | ○ | | | | | | | | ⊙ | | | | |
| | Wave (8.33% of 12 gestures in 8 papers) | ○ | | | | | | | | ⊙ | | | | |
| Create | Defining Area (100% of 30 gestures in 7 pa- pers) | ● | ● | | | | | | | ● | | | | |
| Open Door | Swipe/Slide (66.67% of 6 gestures in 5 pa- pers) | ◉ | | | | | | ● | | | | | | ● |
| | Open Hand (33.33% of 6 gestures in 5 pa- pers) | ⊙ | | | | | | | | | | | | |
| Cut A | Scissor (100% of 6 gestures in 6 pa- pers) | ● | | | | | | | | ● | ● | | ● | |
| Minimize | Close Fingers (60.0% of 6 gestures in 6 pa- pers) | ◉ | ● | | | | | | | | | | | |
| | Swipe/Slide (40.0% of 5 gestures in 4 papers) | ⊙ | | | | | | | | | | | | |
| Maximize | Splay Out Fingers (40.0% of 5 gestures in 4 papers) | ◉ | | | | | | | | | ● | | | |
| | Move One Hand Away from Other Hand (20.0% of 5 gestures in 4 papers) | ○ | | | | | | | | | | | | |
| | Swipe/Slide (20.0% of 5 gestures in 4 papers) | ○ | | | | | | | | | | | | |
| Ungroup | Move Hands Apart (100% of 3 gestures in 3 papers) | ● | ● | | | | | | | ● | ● | | | |

*(continued on next page)*

*(continued from previous page)*

| | | Cross-Domain | 3D modeling | Accessibility and Adaptive System | eLearning and Education | Gaming | HRI | In-Vehicle and Automotive | Large (public) Display | Manipulation/ navigation | Media and Entertainment | Medical Technology | Mobile interaction | Smart Rooms / Appliances |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | Move Hands Together (100% of 3 gestures in 3 papers) | ● | ● | | | | | | | ● | ● | | | |
| Skip | Swipe/Slide (100% of 3 gestures in 3 papers) | ● | | | | | | ● | | | ● | | | |
| Cut B | Slide/Knife (66.67% of 9 gestures in 5 pa- perers) | ◉ | ◉ | | ● | ● | | | | | | | | |
| | Point New Area (11.11% of 9 gestures in 5 pa- pers) | ○ | ○ | | | | | | | | | | | |
| | Karate (11.11% of 9 gestures in 5 pa- pers) | ○ | ○ | | | | | | | | | | | |
| Take Picture | Make Fram (57.14% of 7 gestures in 3 papers) | ◉ | | | | | ● | | | | | | | |
| | Show Palm (28.57% of 7 gestures in 3 papers) | ⊙ | | | | | | | | | | | ● | |
| | Tap (14.28% of 7 gestures in 3 papers) | ○ | | | | | ○ | | | | | | | |
| Move cursor | Move or Slide hand around (77.78% of 18 gestures in 12 papers) | ● | ● | | ● | | | | ● | ● | | | | ● |
| | Pinch/Grab and Move it (11.11% of 18 gestures in 12 papers) | ○ | | | | | | | | | | | | |
| | Rotate arm (5.55% of 18 gestures in 12 papers) | ○ | | | | | | | ● | | | | | |

# Figures

# Tables

# Bibliography


[ACTK15] Ackad, Christopher ; Clayphan, Andrew ; Tomitsch, Martin ; Kay, Judy: An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, S. 1227–1238

[AFT+18] Aromaa, Susanna ; Frangakis, Nikos ; Tedone, Domenico ; Viitaniemi, Juhani ; Aaltonen, Iina: Digital human models in human factors and ergonomics evaluation of gesture interfaces. In: *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), Nr. EICS, S. 1–14

[AFW+18] Aloba, Aishat ; Flores, Gianne ; Woodward, Julia ; Shaw, Alex ; Castonguay, Amanda ; Cuba, Isabella ; Dong, Yuzhu ; Jain, Eakta ; Anthony, Lisa: Kinder-Gator: The UF Kinect Database of Child and Adult Motion. In: *Eurographics (Short Papers)*, 2018, S. 13–16

[AGC+14] Atzori, Manfredo ; Gijsberts, Arjan ; Castellini, Claudio ; Caputo, Barbara ; Hager, Anne-Gabrielle M. ; Elsig, Simone ; Giatsidis, Giorgio ; Bassetto, Franco ; Müller, Henning: Electromyography data for non-invasive naturally-controlled robotic hand prostheses. In: *Scientific data* 1 (2014), Nr. 1, S. 1–13

[AHH+18] Arachchi, SP K. ; Hakim, Noorkholis L. ; Hsu, Hui-Huang ; Klimenko, Stanislav V. ; Shih, Timothy K.: Real-time static and dynamic gesture recognition using mixed space features for 3D virtual world's interactions. In: *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)* IEEE, 2018, S. 627–632

[AHI14] Ahlström, David ; Hasan, Khalad ; Irani, Pourang: Are you comfortable doing that? Acceptance studies of around-device gestures in and for public settings. In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, 2014, S. 193–202

[AHJ+12] Alexander, Jason ; Han, Teng ; Judd, William ; Irani, Pourang ; Subramanian, Sriram: Putting your best foot forward: investigating real-world mappings for foot-based gestures. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, S. 1229–1238

[AJS13] Anastasiou, Dimitra ; Jian, Cui ; Stahl, Christoph: A German-Chinese speech-gesture behavioural corpus of device control in a smart home. In: *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, 2013, S. 1–6

[AKK+19] Arora, Rahul ; Kazi, Rubaiat H. ; Kaufman, Danny M. ; Li, Wilmot ; Singh, Karan: Magicalhands: Mid-air hand gestures for animating in vr. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 2019, S. 463–477

[AKM+20] Ahmed, Naveed ; Kharoub, Hind ; Medjden, Selma ; Alsaafin, Areej ; Lataifeh, Mohammed: A Comparative Study to Assess Human Machine Interaction Mechanisms for Large Scale Displays. In: *2020 IEEE International Conference on Human-Machine Systems (ICHMS)* IEEE, 2020, S. 1–4

[All03] Allen, Gary L.: Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? In: *Spatial cognition and computation* 3 (2003), Nr. 4, S. 259–268

[AMK+15] Adib, Fadel ; Mao, Hongzi ; Kabelac, Zachary ; Katabi, Dina ; Miller, Robert C.: Smart homes that monitor breathing and heart rate. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, S. 837–846

[AMW18] Ali, Abdullah X. ; Morris, Meredith R. ; Wobbrock, Jacob O.: Crowdsourcing similarity judgments for agreement analysis in end-user elicitation studies. In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 2018, S. 177–188

[AMW19] Ali, Abdullah X. ; Morris, Meredith R. ; Wobbrock, Jacob O.: Crowdlicit: A system for conducting distributed end-user elicitation and identification studies. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, S. 1–12

[AQSS16] Amir, Mohd H. ; Quek, Albert ; Sulaiman, Nur Rasyid B. ; See, John: Duke: enhancing virtual reality based fps game with full-body interactions. In: *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, 2016, S. 1–6

[ASDA+19] Alizadeh, Mostafa ; Shaker, George ; De Almeida, Joã. M. ; Morita, Plinio P. ; Safavi-Naeini, Safeddin: Remote monitoring of human vital signs using mm-wave FMCW radar. In: *IEEE Access* 7 (2019), S. 54958–54968

[ASLX+16] Arefin Shimon, Shaikh S. ; Lutton, Courtney ; Xu, Zichun ; Morrison-Smith, Sarah ; Boucher, Christina ; Ruiz, Jaime: Exploring non-touchscreen gestures for smartwatches. In: *Proceedings of the 2016 chi conference on human factors in computing systems*, 2016, S. 3822–3833

[ASW+18] Aslan, Ilhan ; Schmidt, Tabea ; Woehrle, Jens ; Vogel, Lukas ; André, Elisabeth: Pen+ mid-air gestures: Eliciting contextual gestures. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, S. 135–144

[ATBEO13] Al-Timemy, Ali H. ; Bugmann, Guido ; Escudero, Javier ; Outram, Nicholas: Classification of finger movements for the dexterous hand prosthesis control with surface electromyography. In: *IEEE journal of biomedical and health informatics* 17 (2013), Nr. 3, S. 608–618

[AVW13] Anthony, Lisa ; Vatavu, Radu-Daniel ; Wobbrock, Jacob O.: Understanding the consistency of users' pen and finger stroke gesture articulation. In: *Proceedings of Graphics Interface 2013.* 2013, S. 87–94

[AWA20] Aloba, Aishat ; Woodward, Julia ; Anthony, Lisa: FilterJoint: toward an understanding of whole-body gesture articulation. In: *Proceedings of the 2020 International Conference on Multimodal Interaction*, 2020, S. 213–221

[AWB+12] Aigner, Roland ; Wigdor, Daniel ; Benko, Hrvoje ; Haller, Michael ; Lindbauer, David ; Ion, Alexandra ; Zhao, Shengdong ; Koh, JTKV: Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. In: *Microsoft Research TechReport MSR-TR-2012-111* 2 (2012), S. 30

[AWB+14] Adhikarla, Vamsi K. ; Woźniak, Pawel ; Barsi, Attila ; Singhal, Dave ; Kovács, Péter T. ; Balogh, Tibor: Freehand interaction with large-scale 3d map data. In: *2014 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)* IEEE, 2014, S. 1–4

[BAEI16] Badam, Sriram K. ; Amini, Fereshteh ; Elmqvist, Niklas ; Irani, Pourang: Supporting visual exploration for multiple users in large display environments. In: *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)* IEEE, 2016, S. 1–10

[BBC+17] Bostan, Idil ; Buruk, Oğuz Turan ; Canat, Mert ; Tezcan, Mustafa O. ; Yurdakul, Celalettin ; Göksun, Tilbe ; Özcan, Oğuzhan: Hands as a controller: User preferences for hand specific on-skin gestures. In: *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, S. 1123–1134

[BBL93] Baudel, Thomas ; Beaudouin-Lafon, Michel: Charade: remote control of objects using free-hand gestures. In: *Communications of the ACM* 36 (1993), Nr. 7, S. 28–35

[BCL+15] Billinghurst, Mark ; Clark, Adrian ; Lee, Gun u. a.: A survey of augmented reality. In: *Foundations and Trends® in Human–Computer Interaction* 8 (2015), Nr. 2-3, S. 73–272

[BCLJ09] Bott, Jared N. ; Crowley, James G. ; LaViola Jr, Joseph J.: Exploring 3D gestural interfaces for music creation in video games. In: *Proceedings of the 4th international Conference on Foundations of Digital Games*, 2009, S. 18–25

[BF93] Benford, Steve ; Fahlén, Lennart: A spatial model of interaction in large virtual environments. In: *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93* Springer, 1993, S. 109–124

[BFD15] Bullock, Ian M. ; Feix, Thomas ; Dollar, Aaron M.: The Yale human grasping dataset: Grasp, object, and task data in household and machine shop environments. In: *The International Journal of Robotics Research* 34 (2015), Nr. 3, S. 251–255

[BGS+05a] Blank, Moshe ; Gorelick, Lena ; Shechtman, Eli ; Irani, Michal ; Basri, Ronen: Actions as space-time shapes. In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* Bd. 2 IEEE, 2005, S. 1395–1402

[BGS+05b] Blank, Moshe ; Gorelick, Lena ; Shechtman, Eli ; Irani, Michal ; Basri, Ronen: Actions as Space-Time Shapes. In: *The Tenth IEEE International Conference on Computer Vision (ICCV'05)*, 2005, S. 1395–1402

[BH19] Bergström, Joanna ; Hornbæk, Kasper: Human–Computer interaction on the skin. In: *ACM Computing Surveys (CSUR)* 52 (2019), Nr. 4, S. 1–14

[BID+99] Bobick, Aaron F. ; Intille, Stephen S. ; Davis, James W. ; Baird, Freedom ; Pinhanez, Claudio S. ; Campbell, Lee W. ; Ivanov, Yuri A. ; Schütte, Arjan ; Wilson, Andrew: The KidsRoom: A perceptually-based interactive and immersive story environment. In: *Presence* 8 (1999), Nr. 4, S. 369–393

[BKM09] Barrie, Peter ; Komninos, Andreas ; Mandrychenko, Oleksii: A pervasive gesture-driven augmented reality prototype using wireless sensor body area networks. In: *Proceedings of the 6th International Conference on Mobile Technology, Application & Systems*, 2009, S. 1–4

[BKS20] Bhowmick, Shimmila ; Kalita, Pratul ; Sorathia, Keyur: A Gesture Elicitation Study for Selection of Nail Size Objects in a Dense and Occluded Dense HMD-VR. In: *IndiaHCI'20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*, 2020, S. 12–23

[BMR+12] Bailly, Gilles ; Müller, Jörg ; Rohs, Michael ; Wigdor, Daniel ; Kratz, Sven: Shoesense: a new perspective on gestural interaction and wearable applications. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, S. 1239–1248

[Bol80] Bolt, Richard A.: "Put-that-there" Voice and gesture at the graphics interface. In: *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, 1980, S. 262–270

[BRS19] Becker, Vincent ; Rauchenstein, Felix ; Sörös, Gábor: Investigating universal appliance control through wearable augmented reality. In: *Proceedings of the 10th Augmented Human International Conference 2019*, 2019, S. 1–9

[BSN14] Baranwal, Neha ; Singh, Neha ; Nandi, Gora C.: Implementation of MFCC based hand gesture recognition on HOAP-2 using Webots platform. In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* IEEE, 2014, S. 1897–1902

[BW10] Benko, Hrvoje ; Wilson, Andrew D.: Multi-point interactions with immersive omnidirectional visualizations in a dome. In: *ACM International Conference on Interactive Tabletops and Surfaces*, 2010, S. 19–28

[Caf15] Cafaro, Francesco: *Using Framed Guessability to Design Gesture Suites for*
*Embodied Interaction*, University of Illinois Urbana-Champaign, Diss., 2015

[CAMCD+19] Castillo, José C. ; Alonso-Martín, Fernando ; Cáceres-Domínguez, David ; Malfaz, María ; Salichs, Miguel A.: The Influence of Speed and Position in Dynamic Gesture Recognition for Human-Robot Interaction. In: *Journal of Sensors* 2019 (2019), Nr. 1, S. 7060491

[CAT22] Colli Alfaro, Jose G. ; Trejos, Ana L.: User-independent hand gesture recognition classification models using sensor fusion. In: *Sensors* 22 (2022), Nr. 4, S. 1321

[CB15] Chattopadhyay, Debaleena ; Bolchini, Davide: Motor-intuitive interactions based on image schemas: Aligning touchless interaction primitives with human sensorimotor abilities. In: *Interacting with Computers* 27 (2015), Nr. 3, S. 327–343

[CCC16] Chen, Li-Chieh ; Chu, Po-Ying ; Cheng, Yun-Maw: Exploring the ergonomic issues of user-defined mid-air gestures for interactive product exhibition. In: *International Conference on Distributed, Ambient, and Pervasive Interactions* Springer, 2016, S. 180–190

[CCVM17] Coenen, Jorgos ; Claes, Sandy ; Vande Moere, Andrew: The concurrent use of touch and mid-air gestures or floor mat interaction on a public display. In: *Proceedings of the 6th ACM International Symposium on Pervasive Displays*, 2017, S. 1–9

[CDT+22] Cappelle, Hans ; Daronkolaei, Ali G. ; Tsang, Jyh ; Debaillie, Björn ; Ocket, Ilja: Radar-Based Human-Robot Interfaces. In: *Artificial Intelligence for Digitising Industry–Applications*. River Publishers, 2022, S. 221–237

[CEZL15] Cauchard, Jessica R. ; E, Jane L. ; Zhai, Kevin Y. ; Landay, James A.: Drone & me: an exploration into natural human-drone interaction. In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, S. 361–365

[CFM+12] Cuccurullo, Stefania ; Francese, Rita ; Murad, Sharefa ; Passero, Ignazio ; Tucci, Maurizio: A gestural approach to presentation exploiting motion capture metaphors. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, S. 148–155

[CGLA20] Cissé, Karim ; Gandhi, Aprajit ; Lottridge, Danielle ; Amor, Robert: User Elicited Hand Gestures for VR-based Navigation of Architectural Designs. In: *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* IEEE, 2020, S. 1–5

[CGWS15] Cai, Xingquan ; Guo, Tianhang ; Wu, Xin ; Sun, Haiyan: Gesture recognition method based on wireless data glove with sensors. In: *Sensor Letters* 13 (2015), Nr. 2, S. 134–137

[CGX+15] Cheng, Xiaogang ; Ge, Qi ; Xie, Shipeng ; Tang, Guijin ; Li, Haibo: UAV gesture interaction design for volumetric surveillance. In: *Procedia Manufacturing* 3 (2015), S. 6639–6643

[CH18] Chaconas, Nikolas ; Höllerer, Tobias: An evaluation of bimanual gestures on the microsoft hololens. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* IEEE, 2018, S. 1–8

[CHV09] Cooper, Harris ; Hedges, Larry V. ; Valentine, Jeffrey C.: The handbook of research synthesis and meta-analysis 2nd edition. In: *The Hand. of Res. Synthesis and Meta-Analysis, 2nd Ed.* Russell Sage Foundation, 2009, S. 1–615

[CKL+12] Choi, Eunjung ; Kwon, Sunghyuk ; Lee, Donghun ; Lee, Hojin ; Chung, Min K.: Can user-derived gesture be considered as the best gesture for a command?: Focusing on the commands for smart home system. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Bd. 56 SAGE Publications Sage CA: Los Angeles, CA, 2012, S. 1253–1257

[CKLP13] Connell, Sabrina ; Kuo, Pei-Yi ; Liu, Liu ; Piper, Anne M.: A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In: *Proceedings of the 12th International Conference on Interaction Design and Children*, 2013, S. 277–280

[CKS16] Cui, Jian ; Kuijper, Arjan ; Sourin, Alexei: Exploration of natural free-hand interaction for shape modeling using leap motion controller. In: *2016 International Conference on Cyberworlds (CW)* IEEE, 2016, S. 41–48

[CL19] Cao, Rui ; Liu, Yue: Hand controlar: An augmented reality application for learning 3d geometry. In: *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* IEEE, 2019, S. 144–149

[CLA18] Cafaro, Francesco ; Lyons, Leilah ; Antle, Alissa N.: Framed guessability: Improving the discoverability of gestures and body movements for full-body interaction. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, S. 1–12

[CLH+20] Chen, Yu-Chun ; Liao, Chia-Ying ; Hsu, Shuo-wen ; Huang, Da-Yuan ; Chen, Bing-Yu: Exploring User Defined Gestures for Ear-Based Interactions. In: *Proceedings of the ACM on Human-Computer Interaction* 4 (2020), Nr. ISS, S. 1–20

[CLK+14] Cafaro, Francesco ; Lyons, Leilah ; Kang, Raymond ; Radinsky, Josh ; Roberts, Jessica ; Vogt, Kristen: Framed guessability: using embodied allegories to increase user agreement on gesture sets. In: *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, 2014, S. 197–204

[CLYJ25] Cao, Yusheng ; Li, Lingyu ; Yuan, Jiehao ; Jeon, Myounghoon: Head-up Displays Improve Drivers' Performance and Subjective Perceptions with the In-Vehicle Gesture Interaction System. In: *International Journal of Human–Computer Interaction* 41 (2025), Nr. 11, S. 6921–6935

[CMPT12] Cohn, Gabe ; Morris, Daniel ; Patel, Shwetak ; Tan, Desney: Humantenna: using the body as an antenna for real-time whole-body interaction. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, S. 1901–1910

[CMYF15] Chen, Sunjie ; Ma, Hongbin ; Yang, Chenguang ; Fu, Mengyin: Hand gesture based robot control system using leap motion. In: *International Conference on Intelligent Robotics and Applications* Springer, 2015, S. 581–591

[CNVV15] Carter, Marcus ; Newn, Joshua ; Velloso, Eduardo ; Vetere, Frank: Remote gaze and gesture tracking on the microsoft kinect: Investigating the role of feedback. In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, 2015, S. 167–176

[COPB12] Cooper, HM ; Ong, Eng-Jon ; Pugeault, Nicolas ; Bowden, Richard: Sign language recognition using sub-units. In: *Journal of Machine Learning Research* 13 (2012), S. 2205–2231

[CPB+16] Christensen, Allan ; Pedersen, Simon A. ; Bjerre, Per ; Pedersen, Andreas K. ; Stuerzlinger, Wolfgang: Transition times for manipulation tasks in hybrid interfaces. In: *International Conference on Human-Computer Interaction* Springer, 2016, S. 138–150

[CPMX07] Chang, Lillian Y. ; Pollard, Nancy S. ; Mitchell, Tom M. ; Xing, Eric P.: Feature selection for grasp recognition from optical markers. In: *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* IEEE, 2007, S. 2944–2950

[CRS17] Caruso, L ; Russo, R ; Savino, S: Microsoft Kinect V2 vision system in a manufacturing application. In: *Robotics and Computer-Integrated Manufacturing* 48 (2017), S. 174–181

[CSS+16] Chan, Edwin ; Seyed, Teddy ; Stuerzlinger, Wolfgang ; Yang, Xing-Dong ; Maurer, Frank: User elicitation on single-hand microgestures. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, S. 3403–3414

[CT17] Cai, Minghao ; Tanaka, Jiro: Trip together: a remote pair sightseeing system supporting gestural communication. In: *Proceedings of the 5th International Conference on Human Agent Interaction*, 2017, S. 317–324

[CTCG17] Carreira, Micael ; Ting, Karine Lan H. ; Csobanka, Petra ; Gonçalves, Daniel: Evaluation of in-air hand gestures interaction for older people. In: *Universal Access in the Information Society* 16 (2017), Nr. 3, S. 561–580

[CXH15] Chatterjee, Ishan ; Xiao, Robert ; Harrison, Chris: Gaze+ gesture: Expressive, precise and targeted free-space interactions. In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015, S. 131–138

[DAE+08] Deller, Matthias ; Agne, Stefan ; Ebert, Achim ; Dengel, Andreas ; Hagen, Hans ; Klein, Bertin ; Bender, Michael ; Bernardin, Tony ; Hamann, Bernd: Managing a document-based information space. In: *Proceedings of the 13th international conference on Intelligent user interfaces*, 2008, S. 119–128

[DDFES15] Dong, Haiwei ; Danesh, Ali ; Figueroa, Nadia ; El Saddik, Abdulmotaleb: An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. In: *IEEE access* 3 (2015), S. 543–555

[DFA15] Dingler, Tilman ; Funk, Markus ; Alt, Florian: Interaction proxemics: Combining physical spaces for seamless gesture interaction. In: *Proceedings of the 4th International Symposium on Pervasive Displays*, 2015, S. 107–114

[DFGS19] Detjen, Henrik ; Faltaous, Sarah ; Geisler, Stefan ; Schneegass, Stefan: User-defined voice and mid-air gesture commands for maneuver-based interventions in automated vehicles. In: *Proceedings of Mensch und Computer 2019.* 2019, S. 341–348

[DGKL19] Dwivedi, Anany ; Gorjup, Gal ; Kwon, Yongje ; Liarokapis, Minas: Combining electromyography and fiducial marker based tracking for intuitive telemanipulation with a robot arm hand system. In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* IEEE, 2019, S. 1–6

[DGS20] Detjen, Henrik ; Geisler, Stefan ; Schneegass, Stefan: Maneuver-based control interventions during automated driving: Comparing touch, voice, and mid-air gestures as input modalities. In: *2020 IEEE international conference on systems, man, and cybernetics (SMC)* IEEE, 2020, S. 3268–3274

[DHR15] Dobbelstein, David ; Hock, Philipp ; Rukzio, Enrico: Belt: An unobtrusive touch input device for head-worn displays. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, S. 2135–2138

[DKK14] Dinh, Dong-Luong ; Kim, Jeong T. ; Kim, Tae-Seong: Hand gesture recognition and interface via a depth imaging sensor for smart home appliances. In: *Energy Procedia* 62 (2014), S. 576–582

[DL13] Datcu, Dragos ; Lukosch, Stephan: Free-hands interaction in augmented reality. In: *Proceedings of the 1st symposium on Spatial user interaction*, 2013, S. 33–40

[DLG+21] Dai, Qingfeng ; Li, Xiangdong ; Geng, Weidong ; Jin, Wenguang ; Liang, Xiubo: CAPG-MYO: a muscle-computer interface supporting user-defined gesture recognition. In: *Proceedings of the 9th international conference on computer and communications management*, 2021, S. 52–58

[DM15] Dicke, Christina ; Müller, Jörg: Evaluating Mid-air List Interaction for Spatial Audio Interfaces. In: *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, 2015, S. 24–33

[DPBK21] Degraen, Donald ; Piovarči, Michal ; Bickel, Bernd ; Krüger, Antonio: Capturing tactile properties of real surfaces for haptic reproduction. In: *The 34th annual ACM symposium on user interface software and technology*, 2021, S. 954–971

[DRSH18a] Dingler, Tilman ; Rzayev, Rufat ; Shirazi, Alireza S. ; Henze, Niels: Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA : Association for Computing Machinery, 2018 (CHI '18). – ISBN 9781450356206, 1–12

[DRSH18b] Dingler, Tilman ; Rzayev, Rufat ; Shirazi, Alireza S. ; Henze, Niels: Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, S. 1–12

[DSS+19] Delamare, William ; Silpasuwanchai, Chaklam ; Sarcar, Sayan ; Shiraki, Toshiaki ; Ren, Xiangshi: On Gesture Combination: An Exploration of a Solution to Augment Gesture Interaction. In: *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, 2019, S. 135–146

[DSSR16] Dim, Nem K. ; Silpasuwanchai, Chaklam ; Sarcar, Sayan ; Ren, Xiangshi: Designing mid-air TV gestures for blind people using user-and choice-based elicitation approaches. In: *Proceedings of the 2016 ACM conference on designing interactive systems*, 2016, S. 204–214

[DVL16] Dev, Kapil ; Villar, Nicolas ; Lau, Manfred: StandUp: understanding body-part and gestural preferences for first-person 3D modeling. In: *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, 2016, S. 97–102

[ED15] Ergüner, Fatih ; Durdu, Pınar O.: Multimodal natural interaction for 3D images. In: *2015 9th International Conference on Application of Information and Communication Technologies (AICT)* IEEE, 2015, S. 305–309

[EE68] Engelbart, Douglas C. ; English, William K.: A research center for augmenting human intellect. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968, S. 395–410

[EGB+13] Escalera, Sergio ; Gonzàlez, Jordi ; Baró, Xavier ; Reyes, Miguel ; Guyon, Isabelle ; Athitsos, Vassilis ; Escalante, Hugo ; Sigal, Leonid ; Argyros, Antonis ; Sminchisescu, Cristian u. a.: Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary. In: *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, S. 365–368

[EQY+18] Ens, Barrett ; Quigley, Aaron ; Yeo, Hui-Shyong ; Irani, Pourang ; Piumsomboon, Thammathip ; Billinghurst, Mark: Counterpoint: exploring mixed-scale gesture interaction for AR applications. In: *Extended abstracts of the 2018 CHI conference on human factors in computing systems*, 2018, S. 1–6

[FAAF15] Fareed, Memon Md Farhan M. ; Akram, Quraishi I. ; Anees, Shaikh Bilal A. ; Fakih, Awab H.: Gesture based wireless single-armed robot in cartesian 3D space using Kinect. In: *2015 Fifth International Conference on Communication Systems and Network Technologies* IEEE, 2015, S. 1210–1215

[FCXW09] Fang, Yikai ; Chai, Xiujuan ; Xu, Lei ; Wang, Kongqiao: Hand tracking and application in map navigation. In: *Proceedings of the First International Conference on Internet Multimedia Computing and Service*, 2009, S. 196–200

[FFC19] Feng, Shichang ; Feng, Zhiquan ; Cao, Liujuan: Many-to-one gesture-to-command flexible mapping approach for smart teaching interface interaction. In: *IEEE Access* 7 (2019), S. 179517–179531

[FFK08] Fogtmann, Maiken H. ; Fritsch, Jonas ; Kortbek, Karen J.: Kinesthetic interaction: revealing the bodily potential in interaction design. In: *Proceedings of the 20th Australasian conference on computer-human interaction: designing for habitus and habitat*, 2008, S. 89–96

[FGS+20] Faleel, Shariff A. ; Gammon, Michael ; Sakamoto, Yumiko ; Menon, Carlo ; Irani, Pourang: User gesture elicitation of common smartphone tasks for hand proximate user interfaces. In: *Proceedings of the 11th Augmented Human International Conference*, 2020, S. 1–8

[FHB+14] Franz, Alfred M. ; Haidegger, Tamas ; Birkfellner, Wolfgang ; Cleary, Kevin ; Peters, Terry M. ; Maier-Hein, Lena: Electromagnetic tracking in medicine—a review of technology, validation, and applications. In: *IEEE transactions on medical imaging* 33 (2014), Nr. 8, S. 1702–1725

[FLW12] Findlater, Leah ; Lee, Ben ; Wobbrock, Jacob: Beyond QWERTY: augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, S. 2679–2682

[FNGA12a] Farhadi-Niaki, Farzin ; GhasemAghaei, Reza ; Arya, Ali: Empirical study of a vision-based depth-sensitive human-computer interaction system. In: *Proceedings of the 10th asia pacific conference on Computer human interaction*, 2012, S. 101–108

[FNGA12b] Farhadi-Niaki, Farzin ; GhasemAghaei, Reza ; Arya, Ali: Empirical Study of a Vision-Based Depth-Sensitive Human-Computer Interaction System. In: *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction*. New York, NY, USA : Association for Computing Machinery, 2012 (APCHI '12). – ISBN 9781450314961, 101–108

[FPDH15] Fetzer, Toni ; Petry, Christian ; Deinzer, Frank ; Huffstadt, Karsten: 3D interaction design: Increasing the stimulus-response correspondence by using stereoscopic vision. In: *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* Bd. 1 IEEE, 2015, S. 1–6

[FPT12] Francese, Rita ; Passero, Ignazio ; Tortora, Genoveffa: Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, S. 116–123

[FRYJ24] Fan, Xuelong ; Rydgård, Johan ; Yang, Liyun ; Johansson, Peter J.: Assessing Forearm Exertion in Manual Tasks with Surface EMG: A Comparative Analysis of Through-Forearm vs. Muscle Specific EMG placements. In: *bioRxiv* (2024), S. 2024–12

[FSAP11] Feuerstack, Sebastian ; Santos Anjo, Mauro dos ; Pizzolato, Ednaldo B.: Model-based design and generation of a gesture-based user interface navigation control. In: *IHC+ CLIHC* Citeseer, 2011, S. 227–231

[FVM12] Freeman, Dustin ; Vennelakanti, Ramadevi ; Madhvanath, Sriganesh: Freehand pose-based gestural interaction: Studies and implications for interface design. In: *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)* IEEE, 2012, S. 1–6

[GAI+16] Groenewald, Celeste ; Anslow, Craig ; Islam, Junayed ; Rooney, Chris ; Passmore, Peter J. ; Wong, BL: Understanding 3D mid-air hand gestures with interactive surfaces and displays: a systematic literature review. (2016)

[GDMB18] Gross, Mikaylah ; Dara, Joe ; Meyer, Christopher ; Bolchini, Davide: Exploring Aural Navigation by Screenless Access. In: *Proceedings of the 15th International Web for All Conference*, 2018, S. 1–10

[Ges24] GestureTek: *GestureTek History – GestureTek Health*. `https://gesturetekhealth.com/about/history`. Version: 2024. – Accessed: 2025-08-30

[Gez08] Gezici, Sinan: A survey on wireless position estimation. In: *Wireless personal communications* 44 (2008), Nr. 3, S. 263–282

[GFS19] Gentile, Vito ; Fundarò, Daniele ; Sorce, Salvatore: Elicitation and evaluation of zoom gestures for touchless interaction with desktop displays. In: *Proceedings of the 8th ACM International Symposium on Pervasive Displays*, 2019, S. 1–7

[GJWT+18] Grunnet-Jepsen, Anders ; Winer, Paul ; Takagi, Aki ; Sweetser, John ; Zhao, Kevin ; Khuong, Tri ; Nie, Dan ; Woodfill, John: Using the Intel® RealSenseTM Depth cameras D4xx in Multi-Camera Configurations. In: *New Technologies Group, Intel Corporation, Rev* 1 (2018)

[GK12] Guesgen, Hans W. ; Kessell, Darren: Gestural control of household appliances for the physically impaired. In: *Twenty-Fifth International FLAIRS Conference*, 2012

[GMB+11] Greenberg, Saul ; Marquardt, Nicolai ; Ballendat, Till ; Diaz-Marino, Rob ; Wang, Miaosen: Proxemic interactions: the new ubicomp? In: *interactions* 18 (2011), Nr. 1, S. 42–50

[GPI+15] Guy, Emilie ; Punpongsanon, Parinya ; Iwai, Daisuke ; Sato, Kosuke ; Boubekeur, Tamy: LazyNav: 3D ground navigation with non-critical body parts. In: *2015 IEEE symposium on 3D user interfaces (3DUI)* IEEE, 2015, S. 43–50

[GPY+17] Gupta, Aakar ; Pietrzak, Thomas ; Yau, Cleon ; Roussel, Nicolas ; Balakrishnan, Ravin: Summon and select: Rapid interaction with interface controls in mid-air. In: *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, 2017, S. 52–61

[GRSW15] Gable, Thomas M. ; Raja, Siddharth R. ; Samuels, Dean P. ; Walker, Bruce N.: Exploring and evaluating the capabilities of Kinect v2 in a driving simulator environment. In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2015, S. 297–304

[GS19] Ganapathi, Priya ; Sorathia, Keyur: Elicitation Study of Body Gestures for Locomotion in HMD-VR Interfaces in a Sitting-Position. In: *Motion, Interaction and Games*. 2019, S. 1–10

[GSR+16] Gentile, Vito ; Sorce, Salvatore ; Russo, Giuseppe ; Pirrone, Dario ; Gentile, Antonio: A Multimodal Fruition Model for Graphical Contents in Ancient Books. In: *Proceedings of the 17th International Conference on Computer Systems and Technologies 2016*, 2016, S. 65–72

[GV13] Garzotto, Franca ; Valoriani, Matteo: Touchless gestural interaction with small displays: a case study. In: *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, 2013, S. 1–10

[GVV18] Gheran, Bogdan-Florin ; Vanderdonckt, Jean ; Vatavu, Radu-Daniel: Gestures for smart rings: empirical results, insights, and design implications. In: *Proceedings of the 2018 Designing Interactive Systems Conference*, 2018, S. 623–635

[Gwe14] Gwet, Kilem L.: *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC, 2014

[Han05] Han, Jefferson Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005, S. 115–118

[HGS+20] Han, Mo ; Günay, Sezen Y. ; Schirner, Gunar ; Padir, Taşkın ; Erdoğmuş, Deniz: HANDS: a multimodal dataset for modeling toward human grasp intent inference in prosthetic hands. In: *Intelligent service robotics* 13 (2020), S. 179–185

[HH18] Hatscher, Benjamin ; Hansen, Christian: Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain. In: *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, S. 145–153

[HHB16] Hoff, Lynn ; Hornecker, Eva ; Bertel, Sven: Modifying gesture elicitation: Do kinaesthetic priming and increased production reduce legacy bias? In: *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2016, S. 86–91

[HIC+23] Hosseini, Masoumehsadat ; Ihmels, Tjado ; Chen, Ziqian ; Koelle, Marion ; Müller, Heiko ; Boll, Susanne: Towards a consensus gesture set: A survey of mid-air gestures in HCI for maximized agreement across domains. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, S. 1–24

[HKG+16] Hosio, Simo ; Kukka, Hannu ; Goncalves, Jorge ; Kostakos, Vassilis ; Ojala, Timo: Toward meaningful engagement with pervasive displays. In: *IEEE Pervasive Computing* 15 (2016), Nr. 3, S. 24–31

[HLB+10] Henze, Niels ; Löcken, Andreas ; Boll, Susanne ; Hesselmann, Tobias ; Pielot, Martin: Free-hand gestures for music playback: deriving gestures with a user-centred process. In: *Proceedings of the 9th international conference on Mobile and Ubiquitous Multimedia*, 2010, S. 1–10

[HLL+16] Hu, Yingbai ; Li, Zhijun ; Li, Guanglin ; Yuan, Peijiang ; Yang, Chenguang ; Song, Rong: Development of sensory-motor fusion-based manipulation and grasping control for a robotic hand-eye system. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47 (2016), Nr. 7, S. 1169–1180

[HMB24] Hosseini, Masoumehsadat ; Mueller, Heiko ; Boll, Susanne: Controlling the Rooms: How People Prefer Using Gestures to Control Their Smart Homes. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, S. 1–18

[HRGMI14] Hincapié-Ramos, Juan D. ; Guo, Xiang ; Moghadasian, Paymahn ; Irani, Pourang: Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, S. 1063–1072

[HTG+12] Hespanhol, Luke ; Tomitsch, Martin ; Grace, Kazjon ; Collins, Anthony ; Kay, Judy: Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In: *Proceedings of the 2012 International Symposium on Pervasive Displays*, 2012, S. 1–6

[HTWH19] Hoffmann, Fabian ; Tyroller, Miriam-Ida ; Wende, Felix ; Henze, Niels: User-defined interaction for smart homes: voice, touch, or mid-air gestures? In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, 2019, S. 1–7

[HVT+19] Hasson, Yana ; Varol, Gül ; Tzionas, Dimitrios ; Kalevatykh, Igor ; Black, Michael J. ; Laptev, Ivan ; Schmid, Cordelia: Learning joint reconstruction of hands and manipulated objects. In: *CVPR*, 2019

[HZKB17] Hessam, Jahani F. ; Zancanaro, Massimo ; Kavakli, Manolya ; Billinghurst, Mark: Towards optimization of mid-air gestures for in-vehicle interactions. In: *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, 2017, S. 126–134

[HZLT20] He, Zhifan ; Zhang, Ruifo ; Liu, Zheng ; Tan, Zhengyu: A User-Defined Gesture Set for Natural Interaction in a Smart Kitchen Environment. In: *2020 13th International Symposium on Computational Intelligence and Design (ISCID)* IEEE, 2020, S. 122–125

[IIGI11] Ionescu, Dan ; Ionescu, Bogdan ; Gadea, Cristian ; Islam, Shahidul: An intelligent gesture interface for controlling TV sets and set-top boxes. In: *2011 6th IEEE international symposium on applied computational intelligence and informatics (SACI)* IEEE, 2011, S. 159–164

[ILBL12] Ishii, Hiroshi ; Lakatos, Dávid ; Bonanni, Leonardo ; Labrune, Jean-Baptiste: Radical atoms: beyond tangible bits, toward transformable materials. In: *interactions* 19 (2012), Nr. 1, S. 38–51

[Int] Intel RealSense: *Intel RealSense Depth Camera D455*. https://www.intelrealsense.com/depth-camera-d455/, . – Accessed: 2025-08-17

[IU97] Ishii, Hiroshi ; Ullmer, Brygg: Tangible bits: towards seamless interfaces between people, bits and atoms. In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, S. 234–241

[IWSB15] Ismair, Simon ; Wagner, Julie ; Selker, Ted ; Butz, Andreas: Mime: Teaching mid-air pose-command mappings. In: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2015, S. 199–206

[JILC17] Jane, LE ; Ilene, LE ; Landay, James A. ; Cauchard, Jessica R.: Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In: *CHI*, 2017, S. 6794–6799

[JJLL15] Ju, Zhaojie ; Ji, Xiaofei ; Li, Jing ; Liu, Honghai: An integrative framework of human hand gesture segmentation for human–robot interaction. In: *IEEE Systems Journal* 11 (2015), Nr. 3, S. 1326–1336

[JKK+20] Jehanzeb, Muhammad ; Khan, Muhammad S. ; Khan, Muhammad A. ; Hassan, Umair ; Mehmood, Majid ; Babar, Muhammad I.: A new 3D Viewer system based on hand gesture recognition for smart interaction. In: *2020 International Conference on Computing and Information Technology (ICCIT-1441)* IEEE, 2020, S. 1–4

[JNCR18] Jurewicz, Katherina A. ; Neyens, David M. ; Catchpole, Ken ; Reeves, Scott T.: Developing a 3D gestural interface for anesthesia-related human-computer interaction tasks using both experts and novices. In: *Human factors* 60 (2018), Nr. 7, S. 992–1007

[JPH22] Jiang, N ; Pradhan, A ; He, J: Gesture recognition and biometrics electroMyogram (GRABMyo). In: *PhysioNet. doi: https://doi.org/10.13026/dzhr-nj10* (2022)

[JSM+14] Jones, Brett ; Sodhi, Rajinder ; Murdock, Michael ; Mehra, Ravish ; Benko, Hrvoje ; Wilson, Andrew ; Ofek, Eyal ; MacIntyre, Blair ; Raghuvanshi, Nikunj ; Shapira, Lior: Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, S. 637–644

[JT12] Jones, Steven E. ; Thiruvathukal, George K.: *Codename revolution: the Nintendo Wii platform*. MIT Press, 2012

[Jun20] Jung, Im Y.: A review of privacy-preserving human and human activity recognition. In: *International Journal on Smart Sensing and Intelligent Systems* 13 (2020), Nr. 1, S. 1–13

[JWZ+14] Ju, Zhaojie ; Wang, Yuehui ; Zeng, Wei ; Cai, Haibin ; Liu, Honghai: A modified EM algorithm for hand gesture segmentation in RGB-D data. In: *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* IEEE, 2014, S. 1736–1742

[K+05] Karam, Maria u. a.: A taxonomy of gestures in human computer interactions. (2005)

[KAC+18] Koelle, Marion ; Ananthanarayan, Swamy ; Czupalla, Simon ; Heuten, Wilko ; Boll, Susanne: Your smart glasses' camera bothers me!

exploring opt-in and opt-out gestures for privacy mediation. In: *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 2018, S. 473–481

[KB15] Köpsel, Anne ; Bubalo, Nikola: Benefiting from legacy bias. In: *interactions* 22 (2015), Nr. 5, S. 44–47

[KBO+18] Koelle, Marion ; Boll, Susanne ; Olsson, Thomas ; Williamson, Julie ; Profita, Halley ; Kane, Shaun ; Mitchell, Robb: (Un) Acceptable!?! Re-thinking the Social Acceptability of Emerging Technologies. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, S. 1–8

[KC08] Kim, Tae-Kyun ; Cipolla, Roberto: Canonical correlation analysis of video volume tensors for action categorization and detection. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2008), Nr. 8, S. 1415–1428

[KCIO+03] Kuo, Li-Chieh ; Cooney III, William P. ; Oyama, Mineo ; Kaufman, Kenton R. ; Su, Fong-Chin ; An, Kai-Nan: Feasibility of using surface markers for assessing motion of the thumb trapeziometacarpal joint. In: *Clinical Biomechanics* 18 (2003), Nr. 6, S. 558–563

[KD16] Koutsabasis, Panayiotis ; Domouzis, Chris K.: Mid-air browsing and selection in image collections. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2016, S. 21–27

[KF04] Kim, Hyosun ; Fellner, Dieter W.: Interaction with hand gesture for a back-projection wall. In: *Proceedings Computer Graphics International, 2004.* IEEE, 2004, S. 395–402

[KGH85] Krueger, Myron W. ; Gionfriddo, Thomas ; Hinrichsen, Katrin: VIDEOPLACE—an artificial reality. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1985, S. 35–40

[KHI+12] Kim, David ; Hilliges, Otmar ; Izadi, Shahram ; Butler, Alex D. ; Chen, Jiawen ; Oikonomidis, Iason ; Olivier, Patrick: Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In: *Proceedings of the 25th annual ACM symposium on User interface software and technology*, 2012, S. 167–176

[Khu15] Khundam, Chaowanan: First person movement control with palm normal and hand gesture interaction in virtual reality. In: *2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)* IEEE, 2015, S. 325–330

[KJG+11] Kuehne, Hildegard ; Jhuang, Hueihan ; Garrote, Estíbaliz ; Poggio, Tomaso ; Serre, Thomas: HMDB: a large video database for human motion recognition. In: *2011 International conference on computer vision* IEEE, 2011, S. 2556–2563

[KJWL09] Kane, Shaun K. ; Jayant, Chandrika ; Wobbrock, Jacob O. ; Ladner, Richard E.: Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In: *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, 2009, S. 115–122

[KK12] Kishore, PVV ; Kumar, P R.: Segment, track, extract, recognize and convert sign language videos to voice/text. In: *International Journal of Advanced Computer Science and Applications* 3 (2012), Nr. 6

[KKD14] Kollee, Barry ; Kratz, Sven ; Dunnigan, Anthony: Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In: *Proceedings of the 2nd ACM symposium on spatial user interaction*, 2014, S. 40–49

[KKR+14] Kim, Hyoungnyoun ; Kim, Jun-Sik ; Ryu, Kwanghyun ; Cheon, Seyoung ; Oh, Yonghwan ; Park, Ji-Hyung: Task-oriented teleoperation through natural 3D user interaction. In: *2014 11th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)* IEEE, 2014, S. 335–338

[KLCZ12] Kang, Yu ; Li, Zhijun ; Cao, Xiaoqing ; Zhai, Dihua: Robust control of motion/force for robotic manipulators with random time delays. In: *IEEE Transactions on Control Systems Technology* 21 (2012), Nr. 5, S. 1708–1718

[KLK+16] Kerber, Frederic ; Löchtefeld, Markus ; Krüger, Antonio ; McIntosh, Jess ; McNeill, Charlie ; Fraser, Mike: Understanding same-side interactions with wrist-worn devices. In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 2016, S. 1–10

[Kor08] Kortum, Philip: *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces*. Elsevier, 2008

[KPWL16] Katsuragawa, Keiko ; Pietroszek, Krzysztof ; Wallace, James R. ; Lank, Edward: Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2016, S. 128–135

[KRG+12] Kratz, Sven ; Rohs, Michael ; Guse, Dennis ; Müller, Jörg ; Bailly, Gilles ; Nischt, Michael: PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In: *Proceedings of the international working conference on advanced visual interfaces*, 2012, S. 181–188

[KS99] Kumar, Senthil ; Segen, Jakub: Gesture based 3d man-machine interaction using a single camera. In: *Proceedings IEEE International Conference on Multimedia Computing and Systems* Bd. 1 IEEE, 1999, S. 630–635

[Ks05] Karam, Maria ; schraefel, mc: A study on the use of semaphoric gestures to support secondary task interactions. In: *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 2005, S. 1961–1964

[KSJ+20] Kim, Myeongseop ; Seong, Eunjin ; Jwa, Younkyung ; Lee, Jieun ; Kim, Seungjun: A cascaded multimodal natural user interface to reduce driver distraction. In: *IEEE Access* 8 (2020), S. 112969–112984

[KSP20] Kang, Hyo J. ; Shin, Jung-hye ; Ponto, Kevin: A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In: *2020 IEEE conference on virtual reality and 3D user interfaces (VR)* IEEE, 2020, S. 275–284

[KV19] Koutsabasis, Panayiotis ; Vogiatzidakis, Panagiotis: Empirical research in mid-air interaction: A systematic review. In: *International Journal of Human–Computer Interaction* 35 (2019), Nr. 18, S. 1747–1768

[KWH+11] Kühnel, Christine ; Westermann, Tilo ; Hemmert, Fabian ; Kratz, Sven ; Müller, Alexander ; Möller, Sebastian: I'm home: Defining and evaluating a gesture set for smart-home control. In: *International Journal of Human-Computer Studies* 69 (2011), Nr. 11, S. 693–704

[KWM11] Kwapisz, Jennifer R. ; Weiss, Gary M. ; Moore, Samuel A.: Activity recognition using cell phone accelerometers. In: *ACM SigKDD Explorations Newsletter* 12 (2011), Nr. 2, S. 74–82

[LCBO24] Li, Jieqiong ; Coler, Adam S. ; Borhani, Zahra ; Ortega, Francisco R.: Collecting and Analyzing the Mid-Air Gestures Data in Augmented Reality and User Preferences in Closed Elicitation Study. In: *International Conference on Human-Computer Interaction* Springer, 2024, S. 201–215

[LCD+15] Liang, Hui ; Chang, Jian ; Deng, Shujie ; Chen, Can ; Tong, Ruofeng ; Zhang, Jianjun: Exploitation of novel multiplayer gesture-based interaction and virtual puppetry for digital storytelling to develop children's narrative skills. In: *Proceedings of the 14th ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry*, 2015, S. 63–72

[LCG19] Lu, Shufang ; Cai, Li ; Gao, Fei: Immersive interaction design based on perception of vector field climate data. In: *Systems Science & Control Engineering* 7 (2019), Nr. 1, S. 116–124

[LCK+13] Lee, Sang-Su ; Chae, Jeonghun ; Kim, Hyunjeong ; Lim, Youn-kyung ; Lee, Kun-pyo: Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, S. 617–626

[LCZW11] Lu, Zhiyuan ; Chen, Xiang ; Zhao, Zhangyan ; Wang, Kongqiao: A prototype of gesture-based interface. In: *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2011, S. 33–36

[Lee18] Lee, Jeannie S.: Speech and Gestures for Smart-Home Control and Interaction for Older Adults. In: *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, 2018, S. 49–57

[LFJW98] Latoschik, Marc E. ; Frohlich, Martin ; Jung, Bernhard ; Wachsmuth, Ipke: Utilize speech and gestures to realize natural interaction in a virtual environment. In: *IECON'98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No. 98CH36200)* Bd. 4 IEEE, 1998, S. 2028–2033

[LFPP+17] Lopes, Daniel S. ; Figueiredo Parreira, Pedro D. ; Paulo, Soraia F. ; Nunes, Vitor ; Rego, Paulo A. ; Neves, Manuel C. ; Rodrigues, Pedro S. ; Jorge, Joaquim A.: On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface. In: *Journal of biomedical informatics* 72 (2017), S. 140–149

[LFZG15] Long, Lidun ; Fu, Xinsha ; Zhu, Honglei ; Ge, Ting: Finger gesture-based natural user interface for 3D highway alignment design in virtual environment. In: *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)* Bd. 1 IEEE, 2015, S. 105–111

[LGK+16] Lien, Jaime ; Gillian, Nicholas ; Karagozler, M E. ; Amihood, Patrick ; Schwesig, Carsten ; Olson, Erik ; Raja, Hakim ; Poupyrev, Ivan:

Soli: Ubiquitous gesture sensing with millimeter wave radar. In: *ACM Transactions on Graphics (TOG)* 35 (2016), Nr. 4, S. 1–19

[LHY+20] Lu, Yiqin ; Huang, Bingjian ; Yu, Chun ; Liu, Guahong ; Shi, Yuanchun: Designing and Evaluating Hand-to-Hand Gestures with Dual Commodity Wrist-Worn Devices. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (2020), Nr. 1, S. 1–27

[LJJ+17] Liu, Honghai ; Ju, Zhaojie ; Ji, Xiaofei ; Chan, Chee S. ; Khoury, Mehdi: Fuzzy qualitative trigonometry. In: *Human Motion Sensing and Recognition: A Fuzzy Qualitative Approach*. Springer, 2017, S. 35–50

[LKH+13] Lee, Hyong-Euk ; Kang, Nahyup ; Han, Jae-Joon ; Kim, Ji-Yeon ; Kim, Kyung H. ; Kim, James D. ; Kim, ChangYeong: Interactive manipulation and visualization of a deformable 3D organ model for medical diagnostic support. In: *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)* IEEE, 2013, S. 50–55

[LL13] Liao, Hongjian ; Long, Xiaoli: Study on virtual assembly system based on Kinect somatosensory interaction. In: *2013 International Conference on Information Science and Cloud Computing* IEEE, 2013, S. 55–60

[LLD+11] Leithinger, Daniel ; Lakatos, David ; DeVincenzi, Anthony ; Blackshaw, Matthew ; Ishii, Hiroshi: Direct and gestural interaction with relief: a 2.5 D shape display. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, S. 541–548

[LMAV17] Leiva, Luis A. ; Martín-Albo, Daniel ; Vatavu, Radu-Daniel: Synthesizing stroke gestures across user populations: A case for users with visual impairments. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, S. 4182–4193

[LNJ17] Leng, Hoo Y. ; Norowi, Noris M. ; Jantan, Azrul H.: A user-defined gesture set for music interaction in immersive virtual environment. In: *Proceedings of the 3rd International Conference on Human-Computer Interaction and User Experience in Indonesia*, 2017, S. 44–51

[LP14] Luzanin, Ognjan ; Plancak, Miroslav: Hand gesture recognition using low-budget data glove and cluster-trained probabilistic neural network. In: *Assembly Automation* 34 (2014), Nr. 1, S. 94–105

[LRP+19] Lopes, Daniel ; Relvas, Filipe ; Paulo, Soraia ; Rekik, Yosra ; Grisoni, Laurent ; Jorge, Joaquim: FEETICHE: FEET Input for Contactless Hand gEsture Interaction. In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2019, S. 1–10

[LS13] Liu, Lijun ; Shao, Ling: Learning Discriminative Representations from RGB-D Video Data. In: *IJCAI* Bd. 1 Beijing, People's Republic of China, 2013, S. 3

[LSS11] Lacolina, Samuel A. ; Soro, Alessandro ; Scateni, Riccardo: Natural exploration of 3D models. In: *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, 2011, S. 118–121

[LT06] Lozano, Sandra C. ; Tversky, Barbara: *RETRACTED: Communicative gestures facilitate problem solving for both communicators and recipients.* 2006

[LVV18] Lawson, Jean-Yves L. ; Vanderdonckt, Jean ; Vatavu, Radu-Daniel: Mass-computer interaction for thousands of users and beyond. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, S. 1–6

[LVW18] Liu, Zhe ; Vogel, Daniel ; Wallace, James R.: Applying the cumulative fatigue model to interaction on large, multi-touch displays. In: *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, 2018, S. 1–9

[LXS+13] Liu, Hongzhe ; Xi, Yulong ; Song, Wei ; Um, Kyhyun ; Cho, Kyungeun: Gesture-based NUI application for real-time path modification. In: *2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing* IEEE, 2013, S. 446–449

[LZL20] Liu, Jiaxin ; Zhang, Hongxin ; Li, Chuankang: COMTIS: Customizable touchless interaction system for large screen visualization. In: *Virtual Reality & Intelligent Hardware* 2 (2020), Nr. 2, S. 162–174

[Man25] Manus: *Prime 3 Haptic XR*. `https://manus-meta.com/prime-haptic-xr`, 2025. – Accessed: Sep. 2025

[MBGFY19] Mohan, Pallavi ; Boon Goh, Wooi ; Fu, Chi-Wing ; Yeung, Sai-Kit: Head-Fingers-Arms: Physically-Coupled and Decoupled Multimodal Interaction Designs in Mobile VR. In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2019, S. 1–9

[MBS+11] Mrazovac, Bojan ; Bjelica, Milan Z. ; Simić, Djordje ; Tikvić, Srdjan ; Papp, Ištvan: Gesture based hardware interface for RF lighting control. In: *2011 IEEE 9th International Symposium on Intelligent Systems and Informatics* IEEE, 2011, S. 309–314

[MCG+23] Madapana, Naveen ; Chanci, Daniela ; Gonzalez, Glebys ; Zhang, Lingsong ; Wachs, Juan P.: Touchless interfaces in the operating room: a study in gesture preferences. In: *International Journal of Human–Computer Interaction* 39 (2023), Nr. 3, S. 438–448

[McN92] McNeill, David: *Hand and mind: What gestures reveal about thought*. 1992

[MDD+14] Morris, Meredith R. ; Danielescu, Andreea ; Drucker, Steven ; Fisher, Danyel ; Lee, Bongshin ; Schraefel, MC ; Wobbrock, Jacob O.: Reducing legacy bias in gesture elicitation studies. In: *interactions* 21 (2014), Nr. 3, S. 40–45

[MDZ16] Marin, Giulio ; Dominio, Fabio ; Zanuttigh, Pietro: Hand gesture recognition with jointly calibrated leap motion and depth sensor. In: *Multimedia Tools and Applications* 75 (2016), S. 14991–15015

[MFA+14] Mendes, Daniel ; Fonseca, Fernando ; Araujo, Bruno ; Ferreira, Alfredo ; Jorge, Joaquim: Mid-air interactions above stereoscopic interactive tables. In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)* IEEE, 2014, S. 3–10

[MGDS14] Müller, Jörg ; Geier, Matthias ; Dicke, Christina ; Spors, Sascha: The boomroom: mid-air direct interaction with virtual sound sources. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2014, S. 247–256

[MGKP15] Molchanov, Pavlo ; Gupta, Shalini ; Kim, Kihwan ; Pulli, Kari: Short-range FMCW monopulse radar for hand-gesture sensing. In: *2015 IEEE radar conference (RadarCon)* IEEE, 2015, S. 1491–1496

[MGW17] May, Keenan R. ; Gable, Thomas M. ; Walker, Bruce N.: Designing an in-vehicle air gesture set using elicitation methods. In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, S. 74–83

[MHFK07] Mueller, Mary-Rose ; Hill, Linda ; Fontanesi, John ; Kopald, David: Disagreement on immunization recommendations: an analysis of lay-clinician interaction. In: *Journal of Applied Social Science* 1 (2007), Nr. 2, S. 69–76

[MHWC10] Mauney, Dan ; Howarth, Jonathan ; Wirtanen, Andrew ; Capra, Miranda: Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In: *CHi'10 extended abstracts on human factors in computing systems*. 2010, S. 4015–4020

[MK15] Mapari, Rajesh B. ; Kharat, Govind: Real time human pose recognition using leap motion sensor. In: *2015 IEEE international conference on research in computational intelligence and communication networks (ICRCICN)* IEEE, 2015, S. 323–328

[MM09] Mistry, Pranav ; Maes, Pattie: SixthSense: a wearable gestural interface. In: *ACM Siggraph Asia 2009 Sketches*. 2009, S. 1–1

[MMN17] Micucci, Daniela ; Mobilio, Marco ; Napoletano, Paolo: Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. In: *Applied Sciences* 7 (2017), Nr. 10, S. 1101

[MNC+08] Metcalf, Cheryl D. ; Notley, Scott V. ; Chappell, Paul H. ; Burridge, Jane H. ; Yule, Victoria T.: Validation and application of a computational model for wrist and hand movements using surface markers. In: *IEEE Transactions on Biomedical Engineering* 55 (2008), Nr. 3, S. 1199–1210

[MO13] Menelas, Bob-Antoine J. ; Otis, Martin J.: Use of foot for direct interactions with entities of a virtual environment displayed on a mobile device. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics* IEEE, 2013, S. 3745–3750

[Mor12] Morris, Meredith R.: Web on the wall: insights from a multimodal interaction elicitation study. In: *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, 2012, S. 95–104

[MSL16] Malỳ, Ivo ; Sedláček, David ; Leitao, Paulo: Augmented reality experiments with industrial robot in industry 4.0 environment. In: *2016 IEEE 14th international conference on industrial informatics (INDIN)* IEEE, 2016, S. 176–181

[MWW10] Morris, Meredith R. ; Wobbrock, Jacob O. ; Wilson, Andrew D.: Understanding users' preferences for surface gestures. In: *Proceedings of graphics interface 2010*. 2010, S. 261–268

[MYG+16] Molchanov, Pavlo ; Yang, Xiaodong ; Gupta, Shalini ; Kim, Kihwan ; Tyree, Stephen ; Kautz, Jan: Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network.

In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, S. 4207–4215

[MZR+21] Moin, Ali ; Zhou, Andy ; Rahimi, Abbas ; Menon, Alisha ; Benatti, Simone ; Alexandrov, George ; Tamakloe, Senam ; Ting, Jonathan ; Yamamoto, Natasha ; Khan, Yasser u. a.: A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. In: *Nature Electronics* 4 (2021), Nr. 1, S. 54–63

[Mü07] In: *Dynamic Time Warping*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007. – ISBN 978–3–540–74048–3, 69–84

[New94] Newby, Gregory B.: Gesture recognition based upon statistical similarity. In: *Presence: Teleoperators & Virtual Environments* 3 (1994), Nr. 3, S. 236–243

[NFUK99] Nishino, Hiroaki ; Fushimi, Midori ; Utsumiya, Kouichi ; Korida, Kazuyoshi: A virtual environment for modeling 3D objects through spatial interaction. In: *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* Bd. 6 IEEE, 1999, S. 81–86

[NHON14] Nebeling, Michael ; Huber, Alexander ; Ott, David ; Norrie, Moira C.: Web on the wall reloaded: Implementation, replication and refinement of user-defined interaction sets. In: *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 2014, S. 15–24

[NKG14] Nandakumar, Rajalakshmi ; Kellogg, Bryce ; Gollakota, Shyamnath: Wi-fi gesture recognition on existing devices. In: *arXiv preprint arXiv:1411.5394* (2014)

[NKQK13] Nacenta, Miguel A. ; Kamber, Yemliha ; Qiang, Yizhou ; Kristensson, Per O.: Memorability of pre-designed and user-defined gesture sets. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2013, S. 1099–1108

[NLS+11] Nesselrath, Robert ; Lu, Chensheng ; Schulz, Christian H. ; Frey, Jochen ; Alexandersson, Jan: A gesture based system for context–sensitive interaction with smart homes. In: *Ambient Assisted Living: 4. AAL-Kongress 2011, Berlin, Germany, January 25–26, 2011* Springer, 2011, S. 209–219

[NNK16] Nakazato, Kanako ; Nishino, Hiroaki ; Kodama, Toshitada: A desktop 3D modeling system controllable by mid-air interactions. In: *2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)* IEEE, 2016, S. 633–637

[NSMG03] Nielsen, Michael ; Störring, Moritz ; Moeslund, Thomas B. ; Granum, Erik: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: *International gesture workshop* Springer, 2003, S. 409–420

[NSMG04] Nielsen, Michael ; Störring, Moritz ; Moeslund, Thomas B. ; Granum, Erik: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers 5* Springer, 2004, S. 409–420

[NSTVCV16] Nicola, Stelian ; Stoicu-Tivadar, Lăcrămioara ; Virag, Ioan ; Crişan-Vida, Mihaela: Leap motion supporting medical education. In: *2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)* IEEE, 2016, S. 153–156

[OGT+17] Ortega, Francisco R. ; Galvan, Alain ; Tarre, Katherine ; Barreto, Armando ; Rishe, Naphtali ; Bernal, Jonathan ; Balcazar, Ruben ; Thomas, Jason-Lee: Gesture elicitation for 3D travel via multi-touch and mid-Air systems for procedurally generated pseudo-universe. In: *2017 IEEE symposium on 3D User Interfaces (3DUI)* IEEE, 2017, S. 144–153

[OGTCS20] Oliveira, Eder de ; Gonzalez, Esteban W. ; Trevisan, Daniela G. ; Castro Salgado, Luciana C.: Investigating Users' Natural Engagement with a 3D Design Approach in an Egocentric Vision Scenario. In: *2020 22nd Symposium on Virtual and Augmented Reality (SVR)* IEEE, 2020, S. 74–82

[OHFE16] Ouzzani, Mourad ; Hammady, Hossam ; Fedorowicz, Zbys ; Elmagarmid, Ahmed: Rayyan—a web and mobile app for systematic reviews. In: *Systematic Reviews* 5 (2016), Nr. 1, 210. `http://dx.doi.org/10.1186/s13643-016-0384-4`. – DOI 10.1186/s13643–016–0384–4. – ISSN 2046–4053

[OKK+16] Obaid, Mohammad ; Kistler, Felix ; Kasparavičiūtė, Gabrielė ; Yantaç, Asim E. ; Fjeld, Morten: How would you gesture navigate a drone? a user-centered approach to control a drone. In: *Proceedings of the 20th International Academic Mindtrek Conference*, 2016, S. 113–121

[OL12] Ouyang, Tom ; Li, Yang: Bootstrapping personal gesture shortcuts with the wisdom of the crowd and handwriting recognition. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, S. 2895–2904

[ORS12] Oprisescu, Serban ; Rasche, Christoph ; Su, Bochao: Automatic static hand gesture recognition using tof cameras. In: *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* IEEE, 2012, S. 2748–2751

[OTC04] Oshlyansky, Lidia ; Thimbleby, Harold ; Cairns, Paul: Breaking affordance: culture as context. In: *Proceedings of the third Nordic conference on Human-computer interaction*, 2004, S. 81–84

[OTK+19] Ortega, Francisco R. ; Tarre, Katherine ; Kress, Mathew ; Williams, Adam S. ; Barreto, Armando B. ; Rishe, Naphtali D.: Selection and manipulation whole-body gesture elicitation study in virtual reality. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* IEEE, 2019, S. 1723–1728

[OZ00] O'Hagan, Rochelle ; Zelinsky, Alexander: Visual gesture interfaces for virtual environments. In: *Proceedings First Australasian User Interface Conference. AUIC 2000 (Cat. No. PR00515)* IEEE, 2000, S. 73–80

[OZR02] O'Hagan, RG ; Zelinsky, Alexander ; Rougeaux, Sebastien: Visual gesture interfaces for virtual environments. In: *Interacting with Computers* 14 (2002), Nr. 3, S. 231–250

[PBOS10] Perry, Mark ; Beckett, Steve ; O'Hara, Kenton ; Subramanian, Sriram: WaveWindow: public, performative gestural interaction. In: *Acm* *international conference on interactive tabletops and surfaces*, 2010, S. 109–112

[PCBC13] Piumsomboon, Thammathip ; Clark, Adrian ; Billinghurst, Mark ; Cockburn, Andy: User-defined gestures for augmented reality. In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2013, S. 955–960

[PCG+17] Palermo, Francesca ; Cognolato, Matteo ; Gijsberts, Arjan ; Müller, Henning ; Caputo, Barbara ; Atzori, Manfredo: Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data. In: *2017 International Conference on Rehabilitation Robotics (ICORR)* IEEE, 2017, S. 1154–1159

[PGGP13] Pu, Qifan ; Gupta, Sidhant ; Gollakota, Shyamnath ; Patel, Shwetak: Whole-home gesture recognition using wireless signals. In: *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, S. 27–38

[PH17] Peshkova, Ekaterina ; Hitz, Martin: Exploring user-defined gestures to control a group of four UAVs. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* IEEE, 2017, S. 169–174

[PJ19] Poor, G M. ; Jude, Alvin: Interaction can hurt-Exploring gesture-based interaction for users with Chronic Pain. In: *Symposium on Spatial User Interaction*, 2019, S. 1–5

[PJ20] Pons, Patricia ; Jaen, Javier: Interactive spaces for children: gesture elicitation for controlling ground mini-robots. In: *Journal of Ambient Intelligence and Humanized Computing* 11 (2020), Nr. 6, S. 2467–2488

[PKS15] Patro, Anshul ; Kataria, Vineet ; Sharma, Shalini: EducApp. In: *2015 International Conference on Industrial Engineering and Operations Management (IEOM)* IEEE, 2015, S. 1–6

[PLR16] Plaumann, Katrin ; Lehr, David ; Rukzio, Enrico: Who Has the Force?: Solving Conflicts for Multi User Mid-Air Gestures for TVs. In: *TVX*, 2016, S. 25–29

[PMB+21] Page, Matthew J. ; McKenzie, Joanne E. ; Bossuyt, Patrick M. ; Boutron, Isabelle ; Hoffmann, Tammy C. ; Mulrow, Cynthia D. ; Shamseer, Larissa ; Tetzlaff, Jennifer M. ; Akl, Elie A. ; Brennan, Sue E. u. a.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. In: *Bmj* 372 (2021)

[PO15] Pietschmann, Daniel ; Ohler, Peter: Spatial mapping of physical and virtual spaces as an extension of natural mapping: Relevance for interaction design and user experience. In: *Virtual, Augmented and Mixed Reality: 7th International Conference, VAMR 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings 7* Springer, 2015, S. 49–59

[Pol] Polhemus: *Viper16 Electromagnetic Motion Tracking System*. `https://polhemus.com/motion-tracking/all-trackers/viper/`, . – Accessed: 2025-08-17

[PPR+13] Pescarin, Sofia ; Pietroni, Eva ; Rescic, Leonardo ; Wallergård, Mattias ; Omar, Karim ; Rufa, Claudio: NICH: a preliminary theoretical study on Natural Interaction applied to Cultural Heritage contexts. In: *2013 Digital Heritage International Congress (DigitalHeritage)* Bd. 1 IEEE, 2013, S. 355–362

[Pro12] PROSTHETIC KNOWLEDGE: *Kinect Genealogy - A Brief History of Gestural Interfaces*. Rhizome (Rhizome.org Editorial). `https://rhizome.org/editorial/2012/oct/23/prosthetic-knowledge-picks-kinect-genealogy-brief-/`. Version: Oktober 2012. – Accessed: 2025-08-31

[PTWL17] PIETROSZEK, Krzysztof ; TAHAI, Liudmila ; WALLACE, James R. ; LANK, Edward: Watchcasting: Freehand 3D interaction with off-the-shelf smartwatch. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)* IEEE, 2017, S. 172–175

[PVTMV18] PHAM, Tran ; VERMEULEN, Jo ; TANG, Anthony ; MACDONALD VERMEULEN, Lindsay: Scale impacts elicited gestures for manipulating holograms: Implications for ar gesture design. In: *Proceedings of the 2018 Designing Interactive Systems Conference*, 2018, S. 227–240

[QKH+15] QAMAR, Ahmad M. ; KHAN, Ahmed R. ; HUSAIN, Syed O. ; RAHMAN, Mohamed A. ; BASLAMAH, Saleh: A multi-sensory gesture-based occupational therapy environment for controlling home appliances. In: *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, S. 671–674

[QMB+02] QUEK, Francis ; MCNEILL, David ; BRYLL, Robert ; DUNCAN, Susan ; MA, Xin-Feng ; KIRBAS, Cemil ; MCCULLOUGH, Karl E. ; ANSARI, Rashid: Multimodal human discourse: gesture and speech. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 9 (2002), Nr. 3, S. 171–193

[QTGJ15] QUINTERO, Camilo P. ; TATSAMBON, Romeo ; GRIDSETH, Mona ; JÄGERSAND, Martin: Visual pointing gestures for bi-directional human robot interaction in a pick-and-place task. In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* IEEE, 2015, S. 349–354

[RA12] RAUTARAY, Siddharth S. ; AGRAWAL, Anupam: Real time gesture recognition system for interaction in dynamic environment. In: *Procedia Technology* 4 (2012), S. 595–599

[RB10] RICO, Julie ; BREWSTER, Stephen: Usable gestures for mobile interfaces: evaluating social acceptability. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, S. 887–896

[RBM13] RIDUWAN, Mohammad ; BASORI, Ahmad H. ; MOHAMED, Farhan: Finger-based gestural interaction for exploration of 3D heart visualization. In: *Procedia-Social and Behavioral Sciences* 97 (2013), S. 684–690

[RBOF05] REEVES, Stuart ; BENFORD, Steve ; O'MALLEY, Claire ; FRASER, Mike: Designing the spectator experience. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005, S. 741–750

[RCSB14] ROSSOL, Nathaniel ; CHENG, Irene ; SHEN, Rui ; BASU, Anup: Touchfree medical interfaces. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* IEEE, 2014, S. 6597–6600

[RGDA14] RATEAU, Hanae ; GRISONI, Laurent ; DE ARAUJO, Bruno: Mimetic interaction spaces: Controlling distant displays in pervasive environments.

In: *Proceedings of the 19th international conference on Intelligent User Interfaces*, 2014, S. 89–94

[RIAH15] Roy, Kathika ; Idiwal, Durga P. ; Agrawal, Annapurna ; Hazra, Bani: Flex sensor based wearable gloves for robotic gripper control. In: *Proceedings of the 2015 Conference on Advances in Robotics*, 2015, S. 1–5

[RKB+17] Raptis, George E. ; Katsini, Christina ; Belk, Marios ; Fidas, Christos ; Samaras, George ; Avouris, Nikolaos: Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies. In: *proceedings of the 25th conference on user modeling, Adaptation and Personalization*, 2017, S. 164–173

[RKES24] Rastgoo, Razieh ; Kiani, Kourosh ; Escalera, Sergio ; Sabokrou, Mohammad: Multi-modal zero-shot dynamic hand gesture recognition. In: *Expert Systems with Applications* 247 (2024), S. 123349

[RLL11] Ruiz, Jaime ; Li, Yang ; Lank, Edward: User-defined motion gestures for mobile interaction. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, S. 197–206

[RM17] Rodriguez, Isabel B. ; Marquardt, Nicolai: Gesture elicitation study on how to opt-in & opt-out from interactions with public displays. In: *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, 2017, S. 32–41

[RMFK14] Rahman, Md Asif U. ; Miah, Md Saef U. ; Fahad, M A. ; Karmaker, Debajyoti: SHIMPG: Simple human interaction with machine using Physical Gesture. In: *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)* IEEE, 2014, S. 301–305

[RRVL+14] Rovelo Ruiz, Gustavo A. ; Vanacken, Davy ; Luyten, Kris ; Abad, Francisco ; Camahort, Emilio: Multi-viewer gesture-based interaction for omni-directional video. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, S. 4077–4086

[RS13a] Reddy, Kiran V. ; Shah, Mubarak: *UCF50 Action Recognition Dataset*. `http://server.cs.ucf.edu/~vision/data.html`, 2013. – Accessed: 2025-06-19

[RS13b] Reddy, Kishore K. ; Shah, Mubarak: Recognizing 50 human action categories of web videos. In: *Machine vision and applications* 24 (2013), Nr. 5, S. 971–981

[RSES11] Rahman, ASM M. ; Saboune, Jamal ; El Saddik, Abdulmotaleb: Motion-path based in car gesture control of the multimedia devices. In: *Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications*, 2011, S. 69–76

[RVG14] Rekik, Yosra ; Vatavu, Radu-Daniel ; Grisoni, Laurent: Understanding users' perceived difficulty of multi-touch gesture articulation. In: *Proceedings of the 16th International Conference on Multimodal Interaction*, 2014, S. 232–239

[RWP07] Rachovides, Dorothy ; Walkerdine, James ; Phillips, Peter: The conductor interaction method. In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 3 (2007), Nr. 4, S. 1–23

[RYZ11] Ren, Zhou ; Yuan, Junsong ; Zhang, Zhengyou: Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In: *Proceedings of the 19th ACM international conference on Multimedia*, 2011, S. 1093–1096

[S+65] Sutherland, Ivan E. u. a.: The ultimate display. In: *Proceedings of the IFIP Congress* Bd. 2 New York, 1965, S. 506–508

[S+80] Skolnik, Merrill I. u. a.: *Introduction to radar systems*. Bd. 3. McGraw-hill New York, 1980

[SA16] Shaw, Alex ; Anthony, Lisa: Analyzing the articulation features of children's touchscreen gestures. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, S. 333–340

[SAAG00] Starner, Thad ; Auxier, Jake ; Ashbrook, Daniel ; Gandy, Maribeth: The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In: *Digest of Papers. Fourth International Symposium on Wearable Computers* IEEE, 2000, S. 87–94

[Saf08] Saffer, Dan: *Designing gestural interfaces: Touchscreens and interactive devices.* " O'Reilly Media, Inc.", 2008

[SAMK11] Shah, Syed Akhlaq H. ; Ahmed, Ali ; Mahmood, Iftekhar ; Khurshid, Khurram: Hand gesture based user interface for computer using a camera and projector. In: *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* IEEE, 2011, S. 168–173

[SAUK+19] Sun, Yongbin ; Armengol-Urpi, Alexandre ; Kantareddy, Sai Nithin R. ; Siegel, Joshua ; Sarma, Sanjay: Magichand: Interact with iot devices in augmented reality environment. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* IEEE, 2019, S. 1738–1743

[SBCS+12] Seyed, Teddy ; Burns, Chris ; Costa Sousa, Mario ; Maurer, Frank ; Tang, Anthony: Eliciting usable gestures for multi-display environments. In: *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, 2012, S. 41–50

[SBD+15] Stecher, M ; Baseler, E ; Draxler, L ; Fricke, L ; Michel, B ; Zimmermann, A ; Bengler, Klaus: Tracking down the intuitiveness of gesture interaction in the truck domain. In: *Procedia Manufacturing* 3 (2015), S. 3176–3183

[SDNP19] Sumpeno, Surya ; Dharmayasa, I Gede A. ; Nugroho, Supeno Mardi S. ; Purwitasari, Diana: Immersive Hand Gesture for Virtual Museum using Leap Motion Sensor Based on K-Nearest Neighbor. In: *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)* IEEE, 2019, S. 1–6

[SEC+19] Satriadi, Kadek A. ; Ens, Barrett ; Cordeil, Maxime ; Jenny, Bernhard ; Czauderna, Tobias ; Willett, Wesley: Augmented reality map navigation with freehand gestures. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* IEEE, 2019, S. 593–603

[Sen25] SenseGlove: *SenseGlove Nova*. `https://www.senseglove.com/nova/`, 2025. – Accessed: Sep. 2025

[SFC+11] Shotton, Jamie ; Fitzgibbon, Andrew ; Cook, Mat ; Sharp, Toby ; Finocchio, Mark ; Moore, Richard ; Kipman, Alex ; Blake, Andrew: Real-time human pose recognition in parts from single depth images. In: *CVPR 2011* Ieee, 2011, S. 1297–1304

[SGH+12] Song, Peng ; Goh, Wooi B. ; Hutama, William ; Fu, Chi-Wing ; Liu, Xiaopei: A handle bar metaphor for virtual object manipulation with mid-air interaction. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, S. 1297–1306

[Shn81] Shneiderman, Ben: Direct manipulation: A step beyond programming languages. In: *Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems.(Part-II): Human Interface and the User Interface-Volume 1981*, 1981, S. 143

[SHP+96] Sharma, Rajeev ; Huang, Thomas S. ; PavloviC, Vladimir I. ; Zhao, Yunxin ; Lo, Zion ; Chu, Stephen ; Schul, K: Speech/gesture interface to a visual computing environment for molecular biologists. In: *Proceedings of 13th international conference on pattern recognition* Bd. 3 IEEE, 1996, S. 964–968

[SK98] Segen, Jakub ; Kumar, Senthil: Fast and accurate 3D gesture recognition interface. In: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170)* Bd. 1 IEEE, 1998, S. 86–91

[SKWA18] Sampson, Hannah ; Kelly, Dan ; Wünsche, Burkhard C. ; Amor, Robert: A hand gesture set for navigating and interacting with 3d virtual environments. In: *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)* IEEE, 2018, S. 1–6

[SKWL17] Siddhpuria, Shaishav ; Katsuragawa, Keiko ; Wallace, James R. ; Lank, Edward: Exploring at-your-side gestural interaction for ubiquitous environments. In: *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, S. 1111–1122

[SL18] Stec, Kashmiri ; Larsen, Lars B.: Gestures for controlling a moveable tv. In: *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*, 2018, S. 5–14

[SLC04] Schuldt, Christian ; Laptev, Ivan ; Caputo, Barbara: Recognizing human actions: a local SVM approach. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* Bd. 3 IEEE, 2004, S. 32–36

[SLL+14] Schröder, Stephan ; Loftfield, Nina ; Langmann, Benjamin ; Frank, Klaus ; Reithmeier, Eduard: Contactless operating table control based on 3D image processing. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* IEEE, 2014, S. 388–392

[SLW+14] Shen, Junchen ; Luo, Yanlin ; Wang, Xingce ; Wu, Zhongke ; Zhou, Mingquan: Gpu-based realtime hand gesture interaction and rendering for volume datasets using leap motion. In: *2014 International Conference on Cyberworlds* IEEE, 2014, S. 85–92

[SMBK18] Seuter, Matthias ; Macrillante, Eduardo R. ; Bauer, Gernot ; Kray, Christian: Running with drones: desired services and control gestures. In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, 2018, S. 384–395

[Smi94] Smith, Walter R.: The Newton application architecture. In: *Proceedings of COMPCON'94* IEEE, 1994, S. 156–161

[Smi96] Smith, Alison: Upper Limb Disorders–Time to relax? In: *Physiotherapy* 82 (1996), Nr. 1, S. 31–38

[SNG18] Simão, Miguel ; Neto, Pedro ; Gibaru, Olivier: *UC2018 DualMyo Hand Gesture Dataset.* `http://dx.doi.org/10.5281/zenodo.1320922`. Version: 2018

[Sol25] Solutions, DGTech E.: *DG5 VHand 3.0 Data Glove.* `https://www.dg-tech.it/en/products/dataglove-dg5-vhand-3-0/`, 2025. – Accessed: Sep. 2025

[SPHB08] Schlömer, Thomas ; Poppinga, Benjamin ; Henze, Niels ; Boll, Susanne: Gesture recognition with a Wii controller. In: *Proceedings of the 2nd international conference on Tangible and embedded interaction*, 2008, S. 11–14

[SPHK08] Soutschek, Stefan ; Penne, Jochen ; Hornegger, Joachim ; Kornhuber, Johannes: 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* IEEE, 2008, S. 1–6

[SSC23] Sanders, Nathan ; Sener, Elif ; Chen, Karen B.: Eliciting Ergonomic User-Defined Gestures for Virtual Reality: A Pilot Study. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Bd. 67 SAGE Publications Sage CA: Los Angeles, CA, 2023, S. 1510–1515

[SSP+14] Song, Jie ; Sörös, Gábor ; Pece, Fabrizio ; Fanello, Sean R. ; Izadi, Shahram ; Keskin, Cem ; Hilliges, Otmar: In-air gestures around unmodified mobile devices. In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, S. 319–329

[SSTO13] Sabir, Kenneth ; Stolte, Christian ; Tabor, Bruce ; O'Donoghue, Seán I: The Molecular Control Toolkit: Controlling 3D molecular graphics via gesture and voice. In: *2013 IEEE Symposium on Biological Data Visualization (BioVis)* IEEE, 2013, S. 49–56

[Sut64] Sutherland, Ivan E.: Sketch pad a man-machine graphical communication system. In: *Proceedings of the SHARE design automation workshop*, 1964, S. 6–329

[SVA+18] Sharma, Sumita ; Varkey, Blessin ; Achary, Krishnaveni ; Hakulinen, Jaakko ; Turunen, Markku ; Heimonen, Tomi ; Srivastava, Saurabh ; Rajput, Nitendra: Designing gesture-based applications for individuals with developmental disabilities: guidelines from user studies in India. In: *ACM Transactions on Accessible Computing (TACCESS)* 11 (2018), Nr. 1, S. 1–27

[SVAG14] Simeone, Adalberto L. ; Velloso, Eduardo ; Alexander, Jason ; Gellersen, Hans: Feet movement in desktop 3D interaction. In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)* IEEE, 2014, S. 71–74

[SW11] Shiratuddin, Mohd F. ; Wong, Kok W.: Non-contact multi-hand gestures interaction techniques for architectural design in a virtual environment. In: *ICIMU 2011: Proceedings of the 5th international Conference on Information Technology & Multimedia* IEEE, 2011, S. 1–6

[Sys25] SYSTEMS, CyberGlove: *CyberGlove III*. https://www.cyberglovesystems.com/cyberglove-iii, 2025. – Accessed: Sep. 2025

[TAM+16] TSAROUCHI, Panagiota ; ATHANASATOS, Athanasios ; MAKRIS, Sotiris ; CHATZIGEORGIOU, Xenofon ; CHRYSSOLOURIS, George: High level robot programming using body and hand gestures. In: *Procedia Cirp* 55 (2016), S. 1–5

[TCS+21] TSANG, Ing J. ; CORRADI, Federico ; SIFALAKIS, Manolis ; VAN LEEKWIJCK, Werner ; LATRÉ, Steven: Radar-based hand gesture recognition using spiking neural networks. In: *Electronics* 10 (2021), Nr. 12, S. 1405

[TGBT20] TAHERI, Omid ; GHORBANI, Nima ; BLACK, Michael J. ; TZIONAS, Dimitrios: GRAB: A dataset of whole-body human grasping of objects. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16* Springer, 2020, S. 581–600

[Tsa18] TSANDILAS, Theophanis: Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation. In: *ACM Trans. Comput.-Hum. Interact.* 25 (2018), jun, Nr. 3. http://dx.doi.org/10.1145/3182168. – DOI 10.1145/3182168. – ISSN 1073–0516

[TTSI+13] TUNTAKURN, Apivan ; THONGVIGITMANEE, Saowapak S. ; SA-ING, Vera ; HASEGAWA, Shoichi ; MAKHANOV, Stanislav S.: Natural interactive 3D medical image viewer based on finger and arm gestures. In: *The 6th 2013 Biomedical Engineering International Conference* IEEE, 2013, S. 1–5

[Vaf13] VAFAEI, Fereydoon: Taxonomy of gestures in human computer interaction. (2013)

[Val25] VALKOV, Dimitar: *pint-public (GitLab repository)*. https://umtl-git.dfki.de/dimitar.valkov/pint-public, Juni 2025. – Accessed: 2025-10-30

[Vat12] VATAVU, Radu-Daniel: User-defined gestures for free-hand TV control. In: *Proceedings of the 10th European conference on Interactive tv and video*, 2012, S. 45–48

[Vat13a] VATAVU, Radu-Daniel: A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. In: *Journal of Ambient Intelligence and Smart Environments* 5 (2013), Nr. 2, S. 187–211

[Vat13b] VATAVU, Radu-Daniel: The impact of motion dimensionality and bit cardinality on the design of 3D gesture recognizers. In: *International Journal of Human-Computer Studies* 71 (2013), Nr. 4, S. 387–409

[Vat13c] VATAVU, Radu-Daniel: There's a world outside your TV: exploring interactions beyond the physical TV screen. In: *Proceedings of the 11th european conference on Interactive TV and video*, 2013, S. 143–152

[Vat17] VATAVU, Radu-Daniel: Beyond features for recognition: human-readable measures to understand users' whole-body gesture performance. In: *International Journal of Human–Computer Interaction* 33 (2017), Nr. 9, S. 713–730

[VAW13] VATAVU, Radu-Daniel ; ANTHONY, Lisa ; WOBBROCK, Jacob O.: Relative accuracy measures for stroke gestures. In: *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, S. 279–286

[VB04] VOGEL, Daniel ; BALAKRISHNAN, Ravin: Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology*, 2004, S. 137–146

[VBLVCB22] VÁSCONEZ, Juan P. ; BARONA LÓPEZ, Lorena I. ; VALDIVIESO CARAGUAY, Ángel L. ; BENALCÁZAR, Marco E.: Hand gesture recognition using EMG-IMU signals and deep q-networks. In: *Sensors* 22 (2022), Nr. 24, S. 9613

[VD18] VILLANUEVA, Marcel ; DRÖGEHORN, Olaf: Using Gestures to Interact with Home Automation Systems: A Socio-Technical Study on Motion Capture Technologies for Smart Homes. In: *SEEDS International Conference-Sustainable, Ecological, Engineering Design for Society*, 2018

[VDH+19] VULETIC, Tijana ; DUFFY, Alex ; HAY, Laura ; MCTEAGUE, Chris ; CAMPBELL, Gerard ; GREALY, Madeleine: Systematic literature review of hand gestures used in human computer interaction interfaces. In: *International Journal of Human-Computer Studies* 129 (2019), S. 74–94

[VGO+20] VYSOCKỲ, Aleš ; GRUSHKO, Stefan ; OŠČÁDAL, Petr ; KOT, Tomáš ; BABJAK, Jan ; JÁNOŠ, Rudolf ; SUKOP, Marek ; BOBOVSKỲ, Zdenko: Analysis of precision and stability of hand tracking with leap motion sensor. In: *Sensors* 20 (2020), Nr. 15, S. 4088

[VK18] VOGIATZIDAKIS, Panagiotis ; KOUTSABASIS, Panayiotis: Gesture elicitation studies for mid-air interaction: A review. In: *Multimodal Technologies and Interaction* 2 (2018), Nr. 4, S. 65

[VK19] VOGIATZIDAKIS, Panagiotis ; KOUTSABASIS, Panayiotis: Frame-based elicitation of mid-air gestures for a smart home device ecosystem. In: *Informatics* Bd. 6 MDPI, 2019, S. 23

[VK22] VOGIATZIDAKIS, Panagiotis ; KOUTSABASIS, Panayiotis: 'Address and command': Two-handed mid-air interactions with multiple home devices. In: *International Journal of Human-Computer Studies* 159 (2022), S. 102755

[VMK+19] VANDERDONCKT, Jean ; MAGROFUOCO, Nathan ; KIEFFER, Suzanne ; PÉREZ, Jorge ; RASE, Ysabelle ; ROSELLI, Paolo ; VILLARREAL, Santiago: Head and shoulders gestures: Exploring user-defined gestures with upper body. In: *International Conference on Human-Computer Interaction* Springer, 2019, S. 192–213

[VNVVW20] VILLARREAL-NARVAEZ, Santiago ; VANDERDONCKT, Jean ; VATAVU, Radu-Daniel ; WOBBROCK, Jacob O.: A systematic review of gesture elicitation studies: What can we learn from 216 studies? In: *Proceedings of the 2020 ACM designing interactive systems conference*, 2020, S. 855–872

[VSA+15] VELLOSO, Eduardo ; SCHMIDT, Dominik ; ALEXANDER, Jason ; GELLERSEN, Hans ; BULLING, Andreas: The feet in human–computer interaction: A survey of foot-based interaction. In: *ACM Computing Surveys (CSUR)* 48 (2015), Nr. 2, S. 1–35

[VU22] VATAVU, Radu-Daniel ; UNGUREAN, Ovidiu-Ciprian: Understanding gesture input articulation with upper-body wearables for users with upper-body motor impairments. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, S. 1–16

[VVCG11] Vatavu, Radu-Daniel ; Vogel, Daniel ; Casiez, Géry ; Grisoni, Laurent: Estimating the perceived difficulty of pen gestures. In: *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part II 13* Springer, 2011, S. 89–106

[VW15] Vatavu, Radu-Daniel ; Wobbrock, Jacob O.: Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, S. 1325–1334

[VW22] Vatavu, Radu-Daniel ; Wobbrock, Jacob O.: Clarifying agreement calculations and analysis for end-user elicitation studies. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 29 (2022), Nr. 1, S. 1–70

[VZ14] Vatavu, Radu-Daniel ; Zaiti, Ionut-Alexandru: Leap gestures for TV: insights from an elicitation study. In: *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, 2014, S. 131–138

[WARM05] Wobbrock, Jacob O. ; Aung, Htet H. ; Rothrock, Brandon ; Myers, Brad A.: Maximizing the guessability of symbolic input. In: *CHI'05 extended abstracts on Human Factors in Computing Systems*, 2005, S. 1869–1872

[WB03] Wu, Mike ; Balakrishnan, Ravin: Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In: *Proceedings of the 16th annual ACM symposium on User interface software and technology*, 2003, S. 193–202

[WB19] Weidner, Florian ; Broll, Wolfgang: Interact with your car: a user-elicited gesture set to inform future in-car user interfaces. In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, 2019, S. 1–12

[WBM13] Walter, Robert ; Bailly, Gilles ; Müller, Jörg: StrikeAPose: revealing mid-air gestures on public displays. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, S. 841–850

[WBRF13] Weichert, Frank ; Bachmann, Daniel ; Rudak, Bartholomäus ; Fisseler, Denis: Analysis of the accuracy and robustness of the leap motion controller. In: *Sensors* 13 (2013), Nr. 5, S. 6380–6393

[WBVM14] Walter, Robert ; Bailly, Gilles ; Valkanova, Nina ; Müller, Jörg: Cuenesics: using mid-air gestures to select items on interactive public displays. In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, 2014, S. 299–308

[WCB11] Williamson, Julie R. ; Crossan, Andrew ; Brewster, Stephen: Multimodal mobile interactions: usability studies in real world settings. In: *Proceedings of the 13th international conference on multimodal interfaces*, 2011, S. 361–368

[Web03] Webb, Andrew R.: *Statistical pattern recognition.* John Wiley & Sons, 2003

[Wel93] Wellner, Pierre: Interacting with paper on the DigitalDesk. In: *Communications of the ACM* 36 (1993), Nr. 7, S. 87–96

[WFSN19] Wagner Filho, Jorge A. ; Stuerzlinger, Wolfgang ; Nedel, Luciana: Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration. In: *IEEE Transactions on Visualization and Computer Graphics* 26 (2019), Nr. 1, S. 514–524

[WGLL04] Wickens, Christopher D. ; Gordon, Sallie E. ; Liu, Yili ; Lee, J: *An introduction to human factors engineering*. Bd. 2. Pearson Prentice Hall Upper Saddle River, NJ, 2004

[WGO20] Williams, Adam S. ; Garcia, Jason ; Ortega, Francisco: Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. In: *IEEE Transactions on Visualization and Computer Graphics* 26 (2020), Nr. 12, S. 3479–3489

[Wil91] Wilson, Paul: *Computer supported cooperative work:: An introduction*. Springer Science & Business Media, 1991

[Wil12] Williamson, Julie R.: *User experience, performance, and social acceptability: usable multimodal mobile interaction*, University of Glasgow, Diss., 2012

[WLC+12] Wang, Jiang ; Liu, Zicheng ; Chorowski, Jan ; Chen, Zhuoyuan ; Wu, Ying: Robust 3d action recognition with random occupancy patterns. In: *European conference on computer vision* Springer, 2012, S. 872–885

[WLC+14] Wang, Yan ; Liu, Jian ; Chen, Yingying ; Gruteser, Marco ; Yang, Jie ; Liu, Hongbo: E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures. In: *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, S. 617–628

[WLK+14] Weichel, Christian ; Lau, Manfred ; Kim, David ; Villar, Nicolas ; Gellersen, Hans W.: MixFab: a mixed-reality environment for personal fabrication. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, S. 3855–3864

[WLLP14] Wan, Qian ; Li, Yiran ; Li, Changzhi ; Pal, Ranadip: Gesture recognition for smart home applications using portable radar sensors. In: *2014 36th annual international conference of the IEEE engineering in medicine and biology society* IEEE, 2014, S. 6414–6417

[WLS+15] Wang, Wei ; Liu, Alex X. ; Shahzad, Muhammad ; Ling, Kang ; Lu, Sanglu: Understanding and modeling of wifi signal based human activity recognition. In: *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, S. 65–76

[WMW09] Wobbrock, Jacob O. ; Morris, Meredith R. ; Wilson, Andrew D.: User-defined gestures for surface computing. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, S. 1083–1092

[WRB06] Weinland, Daniel ; Ronfard, Remi ; Boyer, Edmond: Free viewpoint action recognition using motion history volumes. In: *Computer vision and image understanding* 104 (2006), Nr. 2-3, S. 249–257

[WSE+07] Wachs, Juan ; Stern, Helman ; Edan, Yael ; Gillam, Michael ; Feied, Craig ; Smith, Mark ; Handler, Jon: Gestix: a doctor-computer sterile gesture interface for dynamic environments. In: *Soft Computing in Industrial Applications: Recent Trends* Springer, 2007, S. 30–39

[WSR+06] Wu, Mike ; Shen, Chia ; Ryall, Kathy ; Forlines, Clifton ; Balakrishnan, Ravin: Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In: *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)* IEEE, 2006, S. 8–pp

[WVM+16] Wozniak, Peter ; Vauderwange, Oliver ; Mandal, Avikarsha ; Javahiraly, Nicolas ; Curticapean, Dan: Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality. In: *Optics Education and Outreach IV* Bd. 9946 SPIE, 2016, S. 234–245

[WW11] Wigdor, Daniel ; Wixon, Dennis: *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011

[WW12] Wu, Huiyue ; Wang, Jianmin: User-defined body gestures for TV-based applications. In: *2012 Fourth International Conference on Digital Home* IEEE, 2012, S. 415–420

[WWF+22] Wang, Yong ; Wang, Di ; Fu, Yunhai ; Yao, Dengke ; Xie, Liangbo ; Zhou, Mu: Multi-hand gesture recognition using automotive FMCW radar sensor. In: *Remote Sensing* 14 (2022), Nr. 10, S. 2374

[WWZ16] Wu, Huiyue ; Wang, Jianmin ; Zhang, Xiaolong: User-centered gesture development in TV viewing environment. In: *Multimedia Tools and Applications* 75 (2016), S. 733–760

[WYC+17] Wang, Yu T. ; Yuan, Lu ; Cai, Juan J. ; Wang, Hui ; Zhang, Qin: Interaction design of gesture control for DMS multi-channel system. In: *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)* IEEE, 2017, S. 1524–1528

[WZ15] Wang, Yixin ; Zhang, Wenzeng: Data glove control of robot hand with force telepresence. In: *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)* IEEE, 2015, S. 314–319

[WZM24] Wang, Nanjia ; Zielasko, Daniel ; Maurer, Frank: User Preferences for Interactive 3D Object Transitions in Cross Reality-An Elicitation Study. In: *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*, 2024, S. 1–9

[WZY+24] Wan, Tingjie ; Zhang, Liangyuting ; Yang, Hongyu ; Irani, Pourang ; Yu, Lingyun ; Liang, Hai-Ning: Exploration of foot-based text entry techniques for virtual reality environments. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, S. 1–17

[WZZ+16] Wan, Jun ; Zhao, Yibing ; Zhou, Shuai ; Guyon, Isabelle ; Escalera, Sergio ; Li, Stan Z.: Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, S. 56–64

[XCL+12] Xu, Dan ; Chen, Yen-Lun ; Lin, Chuan ; Kong, Xin ; Wu, Xinyu: Real-time dynamic gesture recognition system based on depth perception for robot navigation. In: *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)* IEEE, 2012, S. 689–694

[XCW+19] Xiao, Yang ; Chen, Jun ; Wang, Yancheng ; Cao, Zhiguo ; Zhou, Joey T. ; Bai, Xiang: Action recognition for depth video using multi-view dynamic images. In: *Information Sciences* 480 (2019), S. 287–304

[XDM+19] Xuan, Li ; Daisong, Guan ; Moli, Zhou ; Jingya, Zhang ; Xingtong, Liu ; Siqi, Li: Comparison on user experience of mid-air gesture interaction and traditional remotes control. In: *Proceedings of the Seventh International Symposium of Chinese CHI*, 2019, S. 16–22

[XGA+22] Xia, Haijun ; Glueck, Michael ; Annett, Michelle ; Wang, Michael ; Wigdor, Daniel: Iteratively Designing Gesture Vocabularies: A Survey and Analysis of Best Practices in the HCI Literature. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 29 (2022), Nr. 4, S. 1–54

[XJX+18] Xue, Yaxu ; Ju, Zhaojie ; Xiang, Kui ; Chen, Jing ; Liu, Honghai: —A review. In: *IEEE Transactions on Cognitive and Developmental Systems* 11 (2018), Nr. 2, S. 162–175

[XYT13] Xiao, Yang ; Yuan, Junsong ; Thalmann, Daniel: Human-virtual human interaction by upper body gesture understanding. In: *Proceedings of the 19th ACM symposium on virtual reality software and technology*, 2013, S. 133–142

[YLZ+12] Yang, Zhong ; Li, Yi ; Zheng, Yang ; Chen, Weidong ; Zheng, Xiaoxiang: An interaction system using mixed hand gestures. In: *Proceedings of the 10th asia pacific conference on Computer human interaction*, 2012, S. 125–132

[YMG+20] Young, Gareth ; Milne, Hamish ; Griffiths, Daniel ; Padfield, Elliot ; Blenkinsopp, Robert ; Georgiou, Orestis: Designing mid-air haptic gesture controlled user interfaces for cars. In: *Proceedings of the ACM on Human-Computer Interaction* 4 (2020), Nr. EICS, S. 1–23

[YPFK+23] Yamagami, Momona ; Portnova-Fahreeva, Alexandra A. ; Kong, Junhan ; Wobbrock, Jacob O. ; Mankoff, Jennifer: How do people with limited movement personalize upper-body gestures? considerations for the design of personalized and accessible gesture interfaces. In: *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, 2023, S. 1–15

[YWF+16] Yu, Ge ; Wang, Yan ; Fu, Taotao ; Liang, Ji ; Guo, Lili: Visuo-Somatosensory Interaction for Space Telescience of Fluid Experiments. In: *2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI)* IEEE Computer Society, 2016, S. 196–200

[YYYS18] Yan, Yukang ; Yu, Chun ; Yi, Xin ; Shi, Yuanchun: Headgesture: Hands-free input approach leveraging head movements for hmd devices. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (2018), Nr. 4, S. 1–23

[Zha12] Zhang, Zhengyou: Microsoft kinect sensor and its effect. In: *IEEE multimedia* 19 (2012), Nr. 2, S. 4–10

[ZLS+17] Zhao, Lijun ; Li, Xiaoyu ; Sun, Zhenye ; Wang, Ke ; Yang, Chenguang: A robot navigation method based on human-robot interaction for 3D environment mapping. In: *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)* IEEE, 2017, S. 409–414

[ZLW19] Zhou, Feifei ; Li, Xiangyu ; Wang, Zhihua: Efficiently user-independent ultrasonic-based gesture recognition algorithm. In: *2019 IEEE SENSORS* IEEE, 2019, S. 1–4

[ZPV15] Zaiţi, Ionuţ-Alexandru ; Pentiuc, Ştefan-Gheorghe ; Vatavu, Radu-Daniel: On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. In: *Personal and Ubiquitous Computing* 19 (2015), S. 821–838

[ZSKG23] Zhao, Yanhua ; Sark, Vladica ; Krstic, Milos ; Grass, Eckhard: Low complexity radar gesture recognition using synthetic training data. In: *Sensors* 23 (2023), Nr. 1, S. 308

[ZZ19] Zhang, Shun ; Zhang, Shizhou: A novel human-3DTV interaction system based on free hand gestures and a touch-based virtual interface. In: *IEEE Access* 7 (2019), S. 165961–165973

[ZZLW12] Zhang, Kai ; Zhai, Ying ; Leong, Hon W. ; Wang, Shengming: An interaction educational computer game framework using hand gesture recognition. In: *Proceedings of the 4th international conference on internet multimedia computing and service*, 2012, S. 219–222