International Journal of Human-Computer Studies 129 (2019) 95–107

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Hover: Exploring cognitive maps and mid-air pointing for television control

Irina Popovici, Ovidiu-Andrei Schipor, Radu-Daniel Vatavu*

MintViz Lab | MANSiD Research Center University Ştefan cel Mare of Suceava 13 Universităţii, Suceava 720229, Romania

ARTICLE INFO

Keywords:
Smart TVs
Air pointing
Deictic gestures
Vicon
Experiment
Pointing accuracy
Recall
Spatial user interfaces

ABSTRACT

We introduce "Hover," a new concept and interface for remote TV control based on air pointing in the users' *personal, peripersonal,* and *extrapersonal space* and implemented with mid-air spatial shortcuts to viewers' preferred TV channels distributed in the room with no representation other than that created in users' own imagination. Hover exploits memory and spatial orientation skills to enable fast access to TV channels with a mere pointing of the hand in mid-air. A controlled experiment with 17 participants and 9 mid-air loci corresponding to shortcuts to television channels revealed an average user recall rate of 80% and individual rates up to 100% for mid-air loci, pointing accuracy offsets of 10 cm, and 99% recognition rate for air pointing actions. We hope that Hover will inspire researchers and practitioners of user interfaces for smart TVs to consider spatial locations and mid-air as active elements to design new interactive experiences for home entertainment environments centered on the smart TV.

1. Introduction

Home entertainment systems, including smart TVs, have recently experienced thriving progress by largely benefiting from advances in display technology (Lu et al., 2017; McGill et al., 2016; Tokuda et al., 2017), human sensing techniques to detect and understand users' actions (Brown et al., 2015; Vatavu, 2015; Vatavu and Mancas, 2014), connected Internet-of-Things (IoT) devices (Rutledge et al., 2016; Yusufov and Kornilov, 2013), including second-screen applications (Geerts et al., 2014; Holz et al., 2015; Neate et al., 2017), and spatial Augmented Reality (sAR) user interfaces (Jones et al., 2014; 2013; Vatavu, 2013b). The ways in which users control television has evolved from the standard remote control to enhanced TV remotes (Bailly et al., 2011; Vatavu, 2012a), voice input (Rao et al., 2017; Renger et al., 2011), gesture-based interfaces (Dim et al., 2016; Vatavu, 2012b; Vatavu and Zaiţi, 2014), and the use of personal devices, such as smartphones and tablets featuring dedicated applications for television TV (Bernhaupt and Pirker, 2014; Bobeth et al., 2014; Lee et al., 2016).

This richness of input modalities and devices has led to creative designs for interfacing television, which the TV industry manufacturers have incorporated into their smart TV products, such as gesture input (Samsung, 2018), voice recognition (Murnane, 2018), and enhanced TV remote controls (LG, 2018). However, no single approach is without shortcomings: physical remote controls can get lost, dirty, break down, need maintenance and battery replacement, or may prove difficult to operate under diminished motor abilities, such as those occurring with age (Mehrotra, 2018); voice commands can negatively affect user experience because of mis-recognition problems (Jiang et al., 2013), especially under background noise interference; second screens, although preferable to the standard TV remote control and free-hand gestures (Bobeth et al., 2014), are exceeded in terms of preference and performance by gesture input detected by enhanced TV remote controls (Cox et al., 2012); while free-hand and whole-body gesture user interfaces need careful design to make sure that the gesture set does not induce fatigue (Hincapié-Ramos et al., 2014; Jang et al., 2017). In this context, there is still room for the community to explore new interaction techniques for effective remote control of smart TVs, such as gesture input performed at the periphery of user attention (Heijboer et al., 2016). Deictic gestures especially, consisting in mere pointing with the hand in mid-air (Cockburn et al., 2011) may provide a good compromise between the familiar pointing performed with the TV remote control (Bailly et al., 2011; Cox et al., 2012; Vatavu, 2013a), yet without its shortcomings (Bernhaupt et al., 2008; Mehrotra, 2018), and more complex gesture user interface designs (Dim et al., 2016; Zaiţi et al., 2015).

In this work, we address this problem from the perspective of spatial user interfaces by proposing and evaluating a novel type of TV remote control with buttons that "hover" in mid-air. Our prototype, "Hover," exploits users' memory and spatial abilities to enable convenient and

* Corresponding author.
E-mail address: radu.vatavu@usm.ro (R.-D. Vatavu).
URL: http://www.eed.usv.ro/~vatavu (R.-D. Vatavu).

https://doi.org/10.1016/j.ijhcs.2019.03.012
Received 17 April 2018; Received in revised form 7 February 2019; Accepted 29 March 2019
Available online 02 April 2019
1071-5819/

fast access to TV features and controls with a mere pointing of the hand in mid-air. For instance, reaching to a specific locus in front of the body could change the TV channel to "Comedy Central;" pointing to another locus to the left and down changes the channel again to "Eurosport," while pointing the hand upwards, at about 20 cm above the head, instructs the TV to turn NetFlix on. Hover transforms the physical space around the user's body into an interactive medium made possible by cognitive maps that are formed, stored into, and retrieved from users' long-term memory. Moreover, Hover is personal, configurable to each user's memory and spatial skills, always-available, and scales easily to multi-viewer contexts to enable shared control over the TV.

The contributions of this work are as follows:

1. We introduce Hover, an interface for remote TV control based on air pointing in the users' personal, peripersonal, and extrapersonal space. Hover implements users' preferred mappings between spatial loci, defined in a reference system centered on the human body, and frequent TV commands, such as shortcuts to preferred or frequently watched TV and Internet video channels.
2. We present an evaluation of Hover in terms of users' memory performance and spatial abilities to recall mid-air loci and point accurately and precisely to those loci without visual feedback. In a controlled experiment involving 17 participants and a high-resolution motion capture system (Vicon Bonita), we found an average precision offset of 6.5 cm (*i.e.*, the consistency of our participants under repeated trials of pointing to the same loci in thin air) and an average accuracy offset of 10 cm (*i.e.*, how far participants were off from the actual intended loci). We also show that Hover can be effectively implemented with recognition rates close to 99% for various loci layouts.

Hover advances the state-of-the-art in remote controlling the smart TV, including recent work in gesture user interfaces for television (Dezfuli et al., 2012; Dim et al., 2016; Vatavu, 2012b; 2015; Vatavu and Zaiţi, 2014), by demonstrating effective use of human memory and spatial orientation abilities for always-available input. We hope that our empirical exploration of memory-based spatial user interfaces for television based on air pointing will inspire researchers and practitioners to consider spatial locations as an active medium to design rich interactive experiences for our future smart home entertainment environments.

2. Related work

We overview in this section interaction techniques and user interfaces designed and implemented for the interactive TV. We also relate to prior work that proposed eyes-free user interfaces for various applications and contexts of use, such as techniques designed to be operated by means of memory associations or proprioception alone. To position our contribution in the right context in this large literature of interactive systems, we focus our discussion on air pointing techniques for spatial target acquisition (Cockburn et al., 2011), interactions designed for the personal (on-body), peripersonal, and extrapersonal space (Guerreiro et al., 2008; Shoemaker et al., 2010; Vatavu, 2017; Yan et al., 2018), including pointing to loci based on memory recall strategies (Perrault et al., 2015; Schipor and Vatavu, 2018), and we discuss interaction techniques designed to be operated with no visual feedback, such as those enabled by imaginary user interfaces (Dezfuli et al., 2012; Gustafson et al., 2010; 2011; Steins et al., 2013).

2.1. Gesture-based user interfaces for the interactive TV

Previous research on effective ways to control the TV set proposed a wide variety of interaction techniques, input modalities, and input devices. This prior work has addressed standard TV functions, such as changing channels or adjusting the audio volume (Vatavu, 2012b; Vatavu and Zaiţi, 2014), but also more demanding features of complex TV systems, such as text entry (Chagas and Furtado, 2013) or input for multi-screen television systems (Vatavu, 2012a; 2013b). In this section, we focus on gesture user interfaces for smart TVs, including deictic gestures in the form of pointing actions performed in mid-air the direction of the TV screen.

Gesture input for television has been implemented with hand-held devices, such as smartphones and tablets, or by using enhanced TV remote designs (Bailly et al., 2011; Bobeth et al., 2014; Vatavu, 2012a; 2013b). For instance, Bailly et al. (2011) explored "gesture-aware" remote controls that embedded sensing electronics to detect users' hand movements. Their concept and prototype were informed by the key observation that users perform pitch and yaw movements anyway when operating the TV remote control. Vatavu (2012a) and Vatavu (2013b) employed an enhanced remote control (*i.e.*, the Nintendo Wii Remote featuring motion sensing capabilities) to enable effective operation of a multi-screen television system. The augmented remote was used to create virtual TV screens, change their layout, control each screen individually, as well as to transition media from one screen to another (Vatavu, 2013b). Bobeth et al. (2014) compared the TV remote control, free-hand gesture input, and a TV-mirroring application on a tablet, and reported that the tablet surpassed the other two input modalities in terms of user performance and preference. Cox et al. (2012) conducted an evaluation of video game controllers and tablet applications to control the TV and reported higher preference and performance for gesture input using game controllers mimicking a TV remote control, such as the Wiimote.

Contactless input devices that detect and track users' movements, such as video cameras, the Microsoft Kinect sensor (Microsoft, 2018), or the Leap Motion controller (LeapMotion, 2018; Vatavu and Zaiţi, 2014), to name just a few, have also been explored to implement gesture user interfaces to replace the classical TV remote control (Vatavu, 2012b; Vatavu and Pentiuc, 2008; Vatavu and Zaiţi, 2014) or to increase its input expressiveness (Bailly et al., 2011; Vatavu, 2015). For example, the "RemoteTouch" system of Choi et al. (2011) employed an optical touchpad to track the position of the user's thumb finger. Vatavu and Zaiţi (2014) used LeapMotion (2018) to record hand poses and finger movements performed to control various functions on smart TVs, such as to change channels, open and close menus, or open the web browser. A study with eighteen participants revealed a 72.8% recall rate and 15.8% false positive recall for free-hand gestures to effect twenty-one functions on the smart TV. The complexity and variability of unconstrained hand and finger gestures resulted in a low agreement rate (.200 on the unit scale) between participants' preferred gesture commands to control various TV functions (Vatavu and Zaiţi, 2014). Prior work has also explored the suitability of mid-air gestures for controlling the TV (Dim et al., 2016; Jeong et al., 2012; Plaumann et al., 2016). For example, Vatavu (2012b) performed the first investigation of users' preferences for TV gesture input and compiled guidelines and recommendations for designing free-hand and whole-body gestures for smart TVs; and Dim et al. (2016) investigated gesture-based input for TV control for blind people and highlighted the importance of a predefined choice list for gesture elicitation studies.

2.2. Spatial and imaginary user interfaces

Spatial user interfaces transform the physical space into an interactive one by means of sensing and visualization technology, such as spatial Augmented Reality. For example, Colley et al. (2014) explored several mobile Augmented Reality scenarios for human memory augmentation, such as revealing passwords or discovering the stories behind various household objects; and Lindlbauer et al. (2016) combined sAR with shape-changing prototypes to extend the design space of appearances and interactions for actuated interfaces. Hartmann and Vogel (2018) examined several mobile phone pointing techniques for sAR and reported raycasting being the fastest technique for high and

distant targets, while direct contact was fastest for targets located in the close proximity of the user.

A distinct category of spatial user interfaces are screen-less, non-visual, and ultra-mobile (Dezfuli et al., 2012; Gustafson et al., 2013; Lin et al., 2013). Such interfaces either replace visual feedback with haptic alternatives (Toyoura et al., 2010; Vatavu et al., 2016) or rely exclusively on users' imagination, spatial memory, and spatial orientation to operate effectively (Gustafson et al., 2011; Perrault et al., 2015; Steins et al., 2013). For instance, the "BioMetal" glove (Toyoura et al., 2010) is a haptic device for rendering contact sensations for sAR applications by converting optical information into vibrotactile feedback. Vatavu et al. (2016) introduced the concept of "digital vibrons," which are invisible, zero-weight digital matter that users can operate freely with their hands. Although invisible, digital vibrons manifest their presence with controlled vibrations localized on the user's fingers. Among the advantages of spatial user interfaces are mobility, ubiquitous availability, and personalization (Gustafson et al., 2010; Rateau et al., 2014; Steins et al., 2013). Moreover, these interfaces are suitable for users with visual impairments by not relying on visual feedback to operate effectively (Toyoura et al., 2010).

"Imaginary interfaces," a distinct class of spatial user interfaces, have been investigated for their appealing characteristics regarding mobility and flexibility (Dezfuli et al., 2012; Gustafson et al., 2010; 2011; 2013; Steins et al., 2013). Gustafson et al. (2010) coined the concept and demonstrated it with a mid-air non-visual interaction technique: once the origin of the imaginary space is defined, users could point and draw freely in that space. Steins et al. (2013) evaluated imaginary interfaces for mimicking input on various devices, such as a steering wheel or a joystick. Gustafson et al. (2011) explored interactions with a smartphone by mapping its home screen to the user's palm; an experiment revealed that participants were able to recall 61% of the home screen icons on their smartphones and to reliably acquire those icons. The same principle was employed for "PalmRC," a system designed to replace the TV remote control with the palm of the hand (Dezfuli et al., 2012). A more detailed investigation (Gustafson et al., 2013) provided a thorough understanding of palm-based interaction by highlighting the relationship between tactile cues in the palm and the imaginary user interface. Imaginary interfaces have also been successfully applied to smart environments with multiple remote displays or involving multiple users (Rateau et al., 2014). Since a virtual or augmented environment may contain remote objects with which the user may wish to interact, various strategies for the ergonomic optimization of mid-air interactions have also been proposed (Montano Murillo et al., 2017).

One important aspect for deploying effective imaginary user interfaces regards the human ability to recall commands without visual feedback. To address the aspect of effective memorization, Perrault et al. (2015) introduced the "Physical Loci" technique to enable users to learn commands easily by leveraging their spatial, object, and semantic memory. Their study revealed a nearly perfect recall rate of 48 items after one week, which demonstrates the practicality of imaginary user interfaces based on users' memory alone. Fruchard et al. (2018) compared an on-body interaction technique, "BodyLoci," to mid-air marking menus in virtual reality by providing users with various types of semantic assistance. Their experimental results showed that basic learning techniques, such as story-making, improved users' recall performance.

2.3. Air pointing, on body, and peripersonal interactions

Spatial interactions between users, digital content, and the physical environment are enabled by various sensing technologies, among which video sensors (Shoemaker et al., 2010; Vatavu, 2017), motion tracking systems (Bolt, 1980; Schipor and Vatavu, 2018; Vogel and Balakrishnan, 2004; Yan et al., 2018), hand-held devices (Hartmann and Vogel, 2018; Li et al., 2009; Vatavu, 2012a), and wearables (Gheran et al., 2018; Haque et al., 2015; Katsuragawa et al., 2016; Popovici and Vatavu, 2018). In their comprehensive study regarding spatial target acquisition with and without visual feedback, Cockburn et al. (2011) explored the design space of air pointing interactions, and listed and discussed five input dimensions (*i.e.*, reference frame, input scale, input degrees of freedom, feedback modality, and feedback content) and six interaction qualities (*i.e.*, learnability for novices, selection speed for experts, accuracy, expressivity, cognitive effort, and comfort) for air pointing. They also compared three techniques: Raycasting, movements in a 3-D volume, and movements across a 2-D plane, with results showing the latter technique to be both rapid and accurate, even in the absence of visual feedback.

Depending on the locations that are pointed to, a delineation of the interactive space enabled by air pointing techniques reveals the *personal space* (*e.g.*, pointing to locations on the body, such as to pockets Vatavu, 2017), the *peripersonal space* (in the immediate vicinity of the body, where objects can be reached and grasped Yan et al., 2018), and the *extrapersonal space*, which extends beyond the limits of direct manipulation (Perrault et al., 2015). Valuable insights to understand how the human brain operates to deal with these delineations comes from neurophysiology and neuropsychology; see Holmes and Spence (2004) that discussed the neural representation of the body and of the space around it, constructed by the human brain to guide the movement of the body through space, and highlighted the integrated representation of visual, somatosensory, and auditory information in the peripersonal space. Previous work on air pointing interaction techniques addressed the on-body, peripersonal, and extrapersonal spaces, as well as transitions between these spaces and combinations thereof (Shoemaker et al., 2010; Vogel and Balakrishnan, 2004).

Regarding the personal space, Guerreiro et al. (2008) examined body-space gestures, referred to as "mnemonical body shortcuts," to improve the efficiency of interactions with mobile devices; Bergstrom-Lehtovirta et al. (2017) looked at the whole body for always-available input and discussed how input devices, such as smartphones, smartwatches, or remote controls, could be mapped to specific loci on the human body; Shoemaker et al. (2010) designed a suit of body-centric interaction techniques for large wall displays, including virtual shadow embodiment, virtual tools stored at physical locations on the user's body, and body-based data storage and control surfaces; Vatavu (2017) introduced "Smart Pockets," a technique for visualizing content on public displays by means of body-referenced gestures to take out digital content from physical pockets and present that content to a nearby surface; and Harrison and Faste (2014) discussed implications of location and touch for designing on-body user interfaces.

In the peripersonal space, Li et al. (2009) introduced "Virtual Shelves," a concept and technique that leverage the position and orientation of a mobile device in front of the users' body as well as users' spatial awareness and kinesthetic memory to trigger shortcuts to various functions for mobile apps, such as accessing bookmarks and navigational buttons for a web browser. Results showed that users could accurately point to seven regions on the "theta" plane and four regions on the "phi" plane by relying solely on their kinesthetic memory. The "AirTouch" system of Lin et al. (2013) enabled control of home appliances with the help of a virtual control panel, which could be re-positioned at various locations and orientations with respect to the user's body. More recently, Yan et al. (2018) examined eyes-free acquisition of targets to improve the interaction experience in Virtual Reality (VR) environments by reducing head movements and changes in attentional focus and, consequently, alleviating the negative effects of fatigue and VR sickness. Their results showed mean spatial offsets that varied, depending on the vertical angle, between 15 cm and 21 cm, but also large user preference (13 out of 16 participants) for eyes-free vs. eyes-engaged target acquisition in virtual environments.

Researchers have also examined air pointing in the extrapersonal space as well as transitions from the personal and peripersonal to the

extrapersonal space. For example, Schipor and Vatavu (2018) presented experimental results regarding users' preferences and memory performance for pinning digital content at the level of an entire room, inside which users could freely move and anchor digital content to any physical object or mid-air location; Vogel and Balakrishnan (2004) presented design principles for transitioning from public to personal and from implicit to explicit interactions with large ambient displays, to which end they identified four interaction zones and phases to describe such transitions, *i.e.*, the ambient display phase, implicit interaction, subtle interaction, and the personal interaction phase; and Perrault et al. (2015) implemented the "Physical Loci" technique by leveraging air pointing to any object or location from a room.

2.4. Summary

Previous work introduced many techniques to control the TV by adding new features and functionalities to the standard TV remote control (Bailly et al., 2011; Vatavu, 2012a) or by replacing it entirely (Dezfuli et al., 2012; Plaumann et al., 2016), among which gesture-based input has received considerable attention within the premises of delivering users with natural and intuitive interaction modalities to operate the functions of the TV (Dezfuli et al., 2012; Dim et al., 2016; Vatavu, 2012a; Zaiţi et al., 2015). Among the wide variety of gesture types to implement such interfaces, we focus in this work on *deictic gestures* in the form of "air pointing" (Cockburn et al., 2011) as promising candidates for TV control due to the simple structure of their underlying movement: the hand points towards the TV screen or to another location in the room to effect a specific function on the TV, such as changing the channel. To this end, we leverage results from the literature on air pointing interaction techniques, previously designed and evaluated for various application domains (Cockburn et al., 2011; Vatavu, 2017; Yan et al., 2018), as well as previous results on effecting generic commands by leveraging users' spatial, object, and semantic memory (Perrault et al., 2015; Schipor and Vatavu, 2018), which we adapt to the specific context of use involving smart TVs. Our interface, Hover, combines such previous approaches to a new application domain with new characteristics, user demands, entertainment needs, and interactive modes, such as lean-back vs. lean-forward (Ali-Hasan and Soto, 2015; Gunther, 2011). In this context, we present the first prototype and empirical results on air pointing and memory-based interfaces applied to the interactive TV, as we hope to draw the attention of the community to the opportunity of adopting such techniques for smart TVs and, thus, to foster developments towards designing effective user interfaces and rich television experiences for viewers.

3. Hover

Hover is a memory-based, imaginary, spatial user interface for television that consists of the following components:

1. *Sensing technology* that detects and reports the user's hand position in mid-air relative to the human body. In this work, we conduct a controlled experiment using a high-resolution motion sensing and tracking system (Vicon) to form a thorough understanding of user performance with Hover.
2. The sensing technology interfaces and communicates with the *TV set* or the home entertainment system to be controlled. In this work, we implemented Hover using YouTube channels that played from a laptop computer connected to a large, 55 inch diagonal smart TV.
3. A *cognitive map* that connects spatial loci in mid-air with various functions of the TV, such as users' preferred or frequently followed channel. Cognitive maps represent internal mental representations of external features or landmarks from the environment. According to Tolman (1948), individuals acquire, decode, and store cues from the environment in order to create and update their internal mental images of the environment itself to assist with goal completion. With Hover, the mapping is personal, created by each user according to their preferences and best strategies to recall spatial loci in mid-air and their associations with television channels. For example, reaching to a specific locus at about 40 cm in front of the body can represent a command to change the TV channel to "Comedy Central," while pointing to the left and down changes the channel to "Eurosport." Spatial locations are represented and stored by Hover as 3-D points. A number of effective strategies adopted by our participants to implement cognitive maps that associate spatial loci and TV functions are discussed next in the article.
4. *Mid-air*. Although obvious, we wish to stress the importance of the physical space surrounding the user that is transformed by Hover into an interactive medium. Inspired by prior work on air-pointing (Cockburn et al., 2011; Perrault et al., 2015; Schipor and Vatavu, 2018; Yan et al., 2018), Hover turns spatial loci into active controls that can be pointed to in order to trigger various functions on the TV with a simple movement of the hand.

4. Experiment

We conducted a controlled experiment to measure and understand users' memory recall and spatial performance with Hover.

4.1. Participants

Seventeen participants (9 male and 8 female) volunteered for our experiment. Participants' ages ranged between 22 and 38 years old ($M = 25.8$, $SD = 4.8$ years). All participants were right handed. To understand participants' spatial abilities, we administered the Cognitrom Assessment System (CAS++) (Cognitrom, 2018), which reported evaluations in the range from *very poor* to *poor, fair, good*, and *very good* spatial abilities. Five participants (29.4%) scored *very good* on the test, six (35.3%) scored *good*, and five (29.4%) scored *fair*; no participants had *very poor* or *poor* spatial abilities, according to the CAS++ evaluation. Ten participants (58.8%) had a technical background (major in Computer Science), while seven were non-technical (major in Educational Sciences). Participants reported watching television, including Internet videos, such as YouTube channels, between 40 min and 5 h on a daily basis ($M = 2.4$, $SD = 1.3$ h).

4.2. Apparatus

We collected 3-D positioning data using a Vicon motion capture system composed of six Bonita B10 cameras (1 Mp resolution and 250 fps for each camera) that tracked a 14 mm-wide marker placed on the participants' index finger. The motion capture system was connected to an Intel Core i7 3.6 GHz PC with 8 GB RAM running Windows 7 and our custom software application (C#,.NET framework) that implemented the experiment design and logged all 3-D positioning data. A chair was placed at about 2.5 m in front of a 55 inch diagonal TV in the middle of a 4 m × 4 m area, inside which the Vicon system had been previously calibrated to track data with precision down to 0.5 mm of translation and 0.5° rotation.[1] We implemented Hover using YouTube channels that played from a laptop computer connected to the TV. However, during the experiment, we disabled the feature of changing the TV channel in response to air pointing movements to prevent biases in participants' subjective perceptions of their recall and accuracy performance (see description next), which we wanted to measure and understand irrespective of their actual, objective performance with Hover, as logged by the Vicon motion tracking system. However, content was played on the TV screen continuously to create the familiar setting of watching television.

[1] https://www.vicon.com/products/archived-products/bonita .

4.3. Task

The experiment consisted of three phases: one *configuration* phase, during which participants placed their preferred TV channels in mid-air, followed by two *recall* phases. For each phase, participants were invited to sit down on the chair in front of the TV set and were briefed with regards to the goal of the respective phase and the tasks they had to perform.

During *configuration*, participants were asked to list their nine most favorite or frequently watched TV channels on a sheet of paper in decreasing order of preference. Once the list was complete, participants were asked to think about suitable locations in the room where they would anchor the channels from their lists. We did not constrain participants about the locations that they could chose, which could be anything, such as physical objects from the room, regions of space, or specific points in mid-air around their body. The only condition was that participants were able to point to those locations precisely and unambiguously, knowing that they would be asked later to recall and point to each channel as accurately as possible. The only instruction provided to participants was *"Please find a suitable location for each channel in the space around you. Try to create meaningful connections between your channels and the locations in space where you place those channels, as this will help you later on to recall the locations, when asked."* After providing the instructions, the experimenter left the room to prevent influencing or distracting the participant in any way, and instructed the participant to call them back when done. Once out of the room, the experimenter started a timer to measure the thinking time needed by the participant to compile their list of preferred channels and to anchor channels in the room. The timer was stopped before returning to the room at the call of the participant. As we found later during data analysis, participants took on average 3.9 min ($SD = 2.6$) to think of suitable locations for their preferred TV channels. Once the experimenter returned to the room, the channel locations were recorded using the Vicon system in the form of 3-D points of the participant's pointed finger, as follows. An infrared marker was attached to the participant's index finger of the dominant hand. Then, the experimenter asked the participant to point to each location in space where channels were placed, and those 3-D locations were stored.

The first *recall* phase took place 30 min after configuration. The second recall took place the following day, after 24 h had passed since configuration. The two recall phases were identical in all aspects, as follows. The experimenter asked the participant to point to the locations in space where they had originally placed each channel during the configuration phase, by specifying both the name of the channel and its position in the ordered list to aid recollection as much as possible, e.g., *"Please show me where you positioned Comedy Central, ranked second on your list of preferred TV channels."* The participants were asked to maintain this position until the location of the marker was recorded by our software application. There was no effect of actually changing the channel on the TV screen and no feedback from the experimenter regarding how well participants recalled the locations of their preferred TV channels to prevent any influence on their subsequent recall trials. The order of channels was randomized and each channel was presented for three times. In total, each participant performed 9 (channels) × 3 (repetitions) × 2 (recall phases) = 54 trials. At the end of each recall phase, participants used a 5-point Likert scale to state how accurately they thought they had recalled and pointed to loci in mid-air.

The configuration and the first recall phase were separated by a 30 min break, during which participants filled in a questionnaire asking for demographic information, participated in an informal interview with the experimenter to debrief their strategies to position and recall channels in mid-air, and took the spatial orientation test.

4.4. Design

The experiment was a within-subject design with the following two independent variables:

1. CHANNEL, ordinal variable with 9 conditions numbered from 1 to 9, where 1 denotes the most preferred or the most frequently followed TV channel.
2. TIME-OF-RECALL represents the moment when participants were asked to recall the locations where they placed their TV channels, with two values: *same-day* and *next-day*. The first recall, corresponding to the *same-day* condition, was solicited 30 min after the participants completed the configuration phase; the second recall, corresponding to the *next-day* condition, was solicited after 24 h had passed since the configuration phase.

The number of conditions for the CHANNEL variable (9) was informed by the upper limit of Miller's "magical" number 7 ± 2 that reflects *"the span of absolute judgment and the span of immediate memory [that] impose severe limitations on the amount of information that we are able to receive, process, and remember"* (Miller, 1956). Obviously, people can remember much more pieces of abstract information by means of chunking and memorization techniques (Perrault et al., 2015) or due to practice, such as when learning to read and write graphical symbols. Prior work even showed that effective memorization of up to 48 spatial loci is possible with very good results (Perrault et al., 2015). However, our participants had to operate with invisible controls placed at invisible and intangible loci in mid-air without any feedback whatsoever, which complicated matters. Therefore, in our experimental design, we decided to use a number of pieces of information that we already knew people can operate effectively even in their working memory (*i.e.*, about 7) and, to push our participants' performance to their limit, we increased this value to the upper Miller's magical number, 9 (Miller, 1956). Moreover, we asked the participants how many of the nine channels were actually followed frequently. Some participants had a difficult time finding nine channels for their list, and the majority of them had less than nine favorite channel, which indicates that nine channels represents a reasonable upper limit for the context of TV control.

4.5. Measures

We computed and evaluated a set of eight measures of user performance with Hover, which we grouped into the following three categories:

A) **Spatial performance measures** evaluate our participants' performance with Hover that reflects in their understanding of the spatial relationships among objects and regions of 3-D space and their memory recall performance to point to specific 3-D loci in mid-air, as follows:
   (1) THINKING-TIME represents the time needed by participants, expressed in minutes, to think of a suitable spatial mapping for their preferred TV channels, knowing that they would have to recall and point to those loci accurately when asked later to do so.
   (2) ACCURACY-OFFSET evaluates how accurately each participant was able to point to the exact mid-air loci where they had originally positioned the channel shortcuts during the configuration phase. We compute ACCURACY-OFFSET per channel and per TIME-OF-RECALL as the Euclidean distance between two 3-D locations reported by the Vicon system tracking the participant's index finger: the reference location ($\mu \in \mathbb{R}^3$) acquired during the configuration phase, and the recalled location ($p \in \mathbb{R}^3$) reported by the Vicon system:

$$\text{ACCURACY-OFFSET}_{i,j} = \left\| \mu_i - p_{i,j} \right\| \quad (1)$$

   where i denotes the locus (1 to 9) and j the trial (3 trials for each TIME-OF-RECALL condition and 6 total trials per participant). We report ACCURACY-OFFSET measurements in centimeters.

(3) PRECISION-OFFSET, measured per participant, channel, and TIME-OF-RECALL, evaluates how consistent each participant was at pointing to the same locus in mid-air during repeated trials. We computed precision as the average Euclidean distance between all of a participant's trials to point to a specific locus in mid-air:

$$\text{PRECISION-OFFSET}_i = \frac{\sum_{j<k} \left\| p_{i,j} - p_{i,k} \right\|}{\sum_{j<k} 1} \quad (2)$$

where i denotes the locus and j and k index all the trials. Three Euclidean distances are computed for each TIME-OF-RECALL condition. Note that precision measurements do not need the actual channels' loci, μ_i, to compute. We report PRECISION-OFFSET in centimeters.

B) **Layout-related measures** report on the specific channel layouts proposed by our participants in mid-air. For all following definitions, we consider that the i-th channel was positioned at coordinates $\mu_i = (x_i, y_i, z_i) \in \mathbb{R}^3$. We compute the following measures:

(4) VOLUME represents the volume of the cuboid that circumscribes all the physical coordinates of mid-air loci for a given participant:

$$\text{VOLUME} = (\max_i x_i - \min_i x_i) \times (\max_i y_i - \min_i y_i) \times (\max_i z_i - \min_i z_i)$$

where i indexes all loci from 1 to 9. We report VOLUME in cubic meters.

(5) PROXIMITY is a measure of the distance computed between the 3-D coordinates of all the pairs of mid-air loci for a given participant:

$$\text{PROXIMITY} = \mathcal{F}\left(\left\{ \left\| \mu_i - \mu_j \right\| \mid 1 \le i < j \le 9 \right\} \right) \quad (3)$$

where $\mathcal{F}$ is a function of $9 \times 8/2 = 36$ Euclidean distances computed between all the pairs of 9 channels. In this work, we are interested in the MIN-PROXIMITY, MAX-PROXIMITY, and AVG-PROXIMITY values, as follows:

$$\text{MIN-PROXIMITY} = \min_{1 \le i < j \le 9} \left\| \mu_i - \mu_j \right\| \quad (4)$$

$$\text{MAX-PROXIMITY} = \max_{1 \le i < j \le 9} \left\| \mu_i - \mu_j \right\| \quad (5)$$

$$\text{AVG-PROXIMITY} = \frac{1}{36} \sum_{1 \le i < j \le 9} \left\| \mu_i - \mu_j \right\| \quad (6)$$

Understanding proximity is important because, as we will report later in the paper, the human precision to recall and point to spatial loci in mid-air is limited, *i.e.*, up to 20 cm, according to our findings. Therefore, any configuration for which two channels are closer to each other than the limits of human precision to point to them accurately is not viable for a practical system. We report PROXIMITY measurements in centimeters.

(6) SPREAD is a measure of how distributed are the channels in space with respect to their centroid:

$$\text{SPREAD} = \frac{1}{n} \sum_i \left\| \mu_i - c \right\| \quad (7)$$

where c is the centroid of the channels' layout:

$$c = \left(\frac{1}{n} \sum_i x_i, \frac{1}{n} \sum_i y_i, \frac{1}{n} \sum_i z_i \right) \quad (8)$$

We report SPREAD values in centimeters.

(C) **Subjective measures of performance**. Next to our objective measures designed to evaluate the accuracy and precision of recalling spatial locations using air pointing movements, we also wanted to understand participants' perceptions of how well they did with Hover. To this end, we employed the following two measures:

(7) PERCEIVED-RECALL represents participants' subjective perceptions of how well they recalled the mid-air loci of all the channels, measured with a 5-point Likert scale with 1 denoting "very difficult to recall" and 5 being "very easy to recall."

(8) PERCEIVED-ACCURACY represents participants' subjective perceptions of how well they thought they had pointed to the mid-air loci of all the channels, measured with a 5-point Likert scale with 1 denoting "not accurate at all" and 5 being "very accurate."

5. Results

We report in this section our participants' performance with air pointing with no information, representation, feedback, or assistance other than imagined or supported by their own memory. We also discuss participants' preferences for layouts of mid-air loci corresponding to TV channels.

5.1. Preferences for anchoring TV channels in mid-air

We start our analysis by exploring participants' strategies and preferences for associating shortcuts to their preferred TV channels via spatial loci in mid-air. Participants spent on average 3.9 min (SD = 2.6) to think of suitable channel layouts. The channel locations, superimposed for all participants, are shown in Fig. 1, from which it is easy to extract the overall parameters of the distribution, such as practical

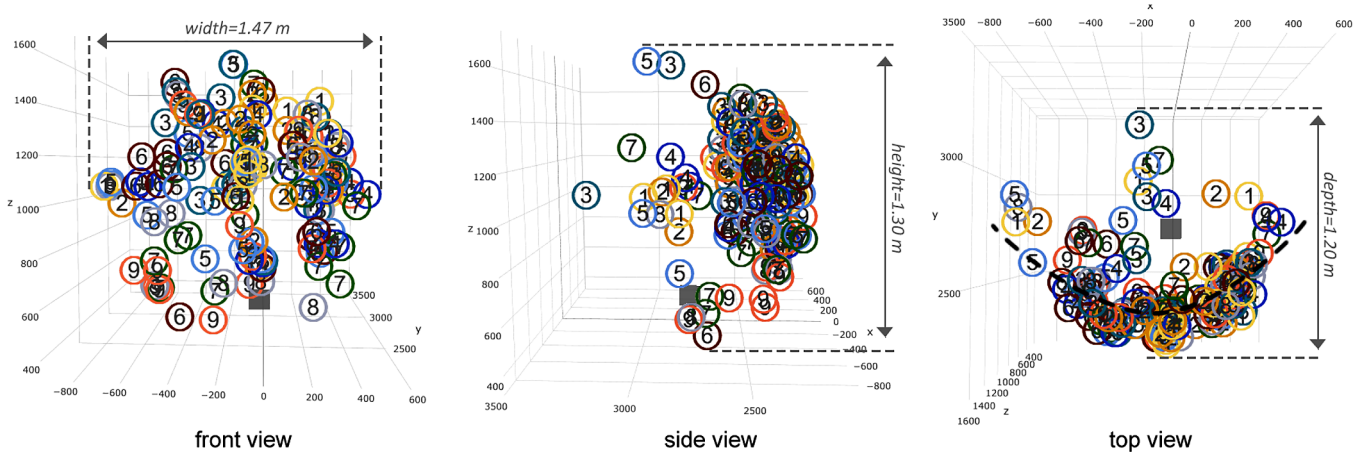

Fig. 1. All layouts superimposed: front view (left), side view (middle), and top view (right). The gray cuboid shows the location of the chair on which participants sat. Note that the majority of channels (89.5%) were anchored in front of the body.

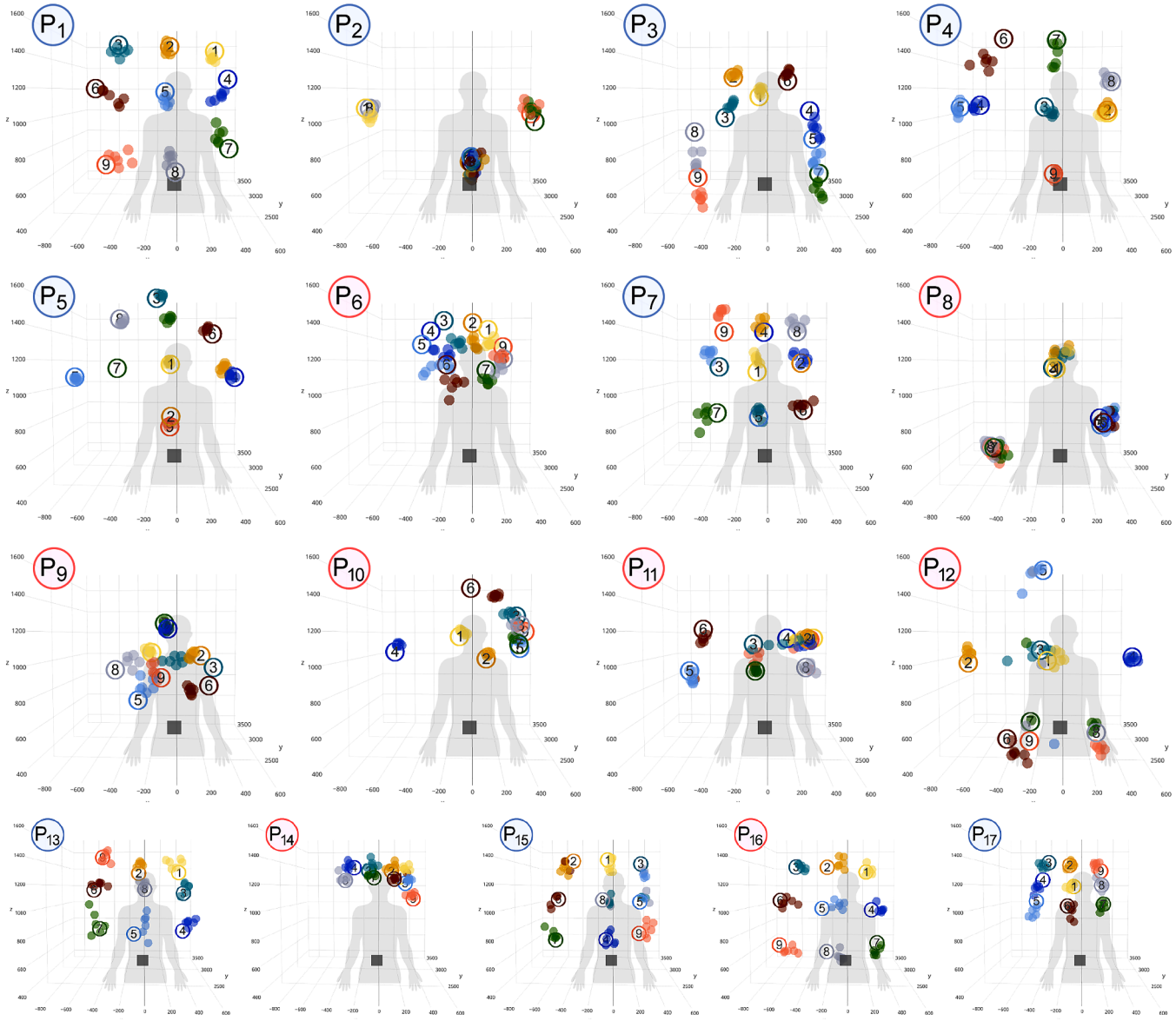

Fig. 2. Layout preferences for each participant. *Notes*: numbers from 1 to 9 encode participants' own choices for TV channels, where 1 denotes the most preferred channel; participants' id and gender are shown in the top-left corner of each layout.

width, height, and depth of the space in which the channels were positioned for efficient reach and recall. Fig. 2 details on all the 918 mid-air loci where our 17 participants pointed to refer to the 9 channels for 6 times repeatedly, illustrating various participants' strategies, such as an easily-identifiable preference for matrix-like layouts (P_1, P_{13} and P_{15}) or forming clusters of TV channels in space (P_2 and P_8).

We found that channels were distributed by our participants in a volume of space centered on their bodies that varied from just 0.02 m^3 to 1.9 m^3 ($M = 0.3$, $SD = 0.5$). MIN-PROXIMITY varied between 1.0 cm and 30.6 cm ($M = 14.2$, $SD = 9.1$), while MAX-PROXIMITY between 61.4 cm and 140.3 cm ($M = 99.8$, $SD = 25.2$). On average, participants positioned channels in mid-air that were approximately at 56 cm one from another ($SD = 15.8$ cm) with an average SPREAD of 40.8 cm ($SD = 11.4$ cm). Table 1 shows a detailed view of the numerical characteristics of all participants' layouts. Out of all the 153 configuration trials (= 17 participants × 9 channels), the large majority (89.5%) were placed in front of the body. Probably due to reasons of hand dominance, our participants placed more channels to their right (56.9%) than to their left (43.1%). Also, 20.9% of the channels were positioned above the head, while only 3.9% were placed lower than the level of the chair seat. The active volume of space circumscribing all these loci was 2.2 m^3, with a range of 1.47 m on the horizontal axis (left-right), 1.30 m on the vertical axis (up-down), and 1.20 m in the front and back of the body; see Fig. 1. The loci of the channels that were chosen in front of the body revealed an interesting pattern, for which we found a statistically significant quadratic regression ($y = 0.092 \cdot x^2 + 0.174 \cdot x - 4.55$ [cm], $R^2 = .398$, $F_{(2,134)} = 44.293$, $p < .001$).[2]

Some layouts are interesting to examine in more detail. For instance, participant P_{12}'s layout had the largest spread (62.1 cm) and volume (1.9 m^3) resulted from using all the available space around their body: front and back, to the right and left, up and down; see Fig. 2. P_2 also proposed a layout with a large spread (55.5 cm), but the volume was much smaller (0.3 m^3), which led to three clusters of channels located in the front, to the left, and to the right of their body. Some participants used a very small volume of space. For example, P_{14}'s

[2] Regression coefficients are expressed in a system of reference centered on the user's body, which we approximated with the center of the chair on which the participants sat during the experiment.

Table 1
Numerical characterization of participants' preferred layouts for TV channels positioned in mid-air, in descending order of the number of effective channels (higher is better). *Note:* higher MIN, AVG, and MAX proximity values, as well as a larger VOLUME and SPREAD denote better-suited layouts.

	Spatial	PROXIMITY (cm)			VOLUME	SPREAD	NUMBER OF CHANNELS		RECOGNITION RATE (%)	
P_{id}	skills score	MIN	MAX	AVG	(m^3)	(cm)	EFFECTIVE[a]	FAVORITE[b]	*same-day*	*next-day*
P_{12}	*fair*	23.6	138.4	88.4	1.91	62.1	9	8	100%	100%
P_4	*very good*	20.7	133.1	77.7	0.74	55.0	9	9	100%	100%
P_1	*fair*	21.0	108.0	66.6	0.19	47.0	9	4	100%	100%
P_{16}	*good*	30.6	107.8	62.7	0.21	44.2	9	6	100%	100%
P_{15}	*very good*	25.9	99.2	54.2	0.09	38.2	9	3	100%	100%
P_7	*good*	23.2	76.7	44.4	0.06	31.3	9	3	93.3%	100%
P_5	*very good*	6.5	129.1	70.9	0.70	50.2	8	4	100%	100%
P_9	*very good*	13.3	96.6	58.4	0.40	40.9	8	4	100%	100%
P_3	*very good*	19.7	95.1	56.8	0.11	41.1	8	6	91.3%	89.5%
P_{13}	*fair*	13.5	93.2	51.8	0.07	36.8	7	5	100%	100%
P_{11}	*good*	3.0	96.5	47.2	0.07	35.9	6	4	91.7%	100%
P_{10}	*good*	9.7	100.2	46.1	0.16	35.1	6	4	100%	100%
P_6	*good*	8.7	61.5	36.1	0.04	25.9	6	3	83.3%	100%
P_{17}	*good*	12.6	61.4	34.7	0.03	24.8	5	4	100%	100%
P_{14}	*fair*	7.2	63.0	28.6	0.02	21.5	4	4	100%	100%
P_2	*very good*	1.0	140.3	67.7	0.26	55.5	3	9	100%	100%
P_8	*fair*	1.0	96.4	60.9	0.25	47.0	3	6	100%	100%
Average		**14.2**	**99.8**	**56.1**	**0.31**	**40.7**	**6.9**	**5.1**	**97.6%**	**99.4%**

[a] The number of effective channels represents the maximum number of channels, out of nine, that are at a distance of at least 20 cm in space, according to our findings on the PROXIMITY measures; see the text for description.
[b] The number of favorite channels represents the channels actually watched on a regular basis by participants out of the nine channels solicited during the configuration phase; see the Experiment section.

layout had a spread of 21.5 cm, but a volume of just 0.02 m^3, which was a consequence of placing all the channels upward by pointing to various regions around the TV screen. P_{17} also placed all the channels in a small volume of just 0.03 m^3 by employing a very narrow matrix-like spatial layout.

The strategies employed by our participants to anchor channels in mid-air can be grouped into five categories, as follows:

1. *Regular spatial layouts.* Seven participants (P_1, P_7, P_{11}, P_{13}, P_{15}, P_{16}, and P_{17}; see Fig. 2) used a matrix layout to anchor channels in mid-air as this specific structure reminded them of the arrangement of buttons on standard TV remote controls or that of the keys on keyboards, such as the T9 layout. Four participants (P_1, P_{11}, P_{13}, and P_{15}) positioned their favorite channel (ranked 1st in their lists and denoted with label "1" in Fig. 2) in the top-left corner of their layouts, whereas the other three participants placed the first channel in the center of the matrix layout for easy access. Two participants did not use the word "matrix" when describing their strategies, but nevertheless ended up using this arrangement. P_6 described the matrix layout as a "square," whereas P_7 created this layout involuntarily by positioning channels in space so that they would be easy to access: in the center, up, down, to the right and left, and in the corners of a vertical imagined surface located in front of the body.
2. *Semantic clustering of channels.* Three participants (P_2, P_4 and P_8) grouped TV channels in space according to their genre. For example, participant P_2 placed all their "tech" channels to the right of the body, the video blog channels to the left, while all the comedy and music related channels went in front of the body. P_4 arranged channels in space at about 45° one from another with respect to the pointing arm, placing the most favorite channels to the right of the body and the least favorite to the left. For most of these layouts, the first channel was positioned either in the front of the body or to the right, where it was easy to reach with the dominant (right) hand.
3. *Anchoring channels to objects in the room.* Five participants (P_3, P_5, P_{10}, P_{11}, and P_{14}) used physical objects located in the room, such as the TV frame, as anchors to position channels. Physical objects can create powerful analogies to help remember information and associations (Perrault et al., 2015). For example, P_{11} pointed towards the power outlet (motivated by frequent daily use), P_{10} pointed to books from a shelf found in the room, while P_3 used the four corners of the TV frame as visible targets to guide their pointing actions.
4. *Using the body as a reference.* We observed preferences for using the body as a physical anchor to refer and point to channels. For example, P_9 placed all the channels on their body: on shoulders, head, waist, hands and knees; see Fig. 2. P_9's two most favorite channels were placed on the two shoulders, whereas the least favorite channel (ranked 9th in the list) was placed at the back of the head. P_{12} distributed all the channels almost equidistantly around the body: the favorite channel in the front, the next channels from the list to the right, in the back, left, and upwards, leaving the least favorite four channels to be placed down near the legs.

5.2. Accuracy and precision of air pointing

We continue our examination by looking at participants' precision and accuracy performance when pointing to spatial loci in order to understand their memory and spatial abilities to use Hover effectively.

A range analysis of the PRECISION-OFFSET measure showed values between 1.2 cm and 95.0 cm for the *same-day* condition (30 min after the configuration phase) and between 1.5 cm and 99.3 cm on the *next day* (after 24 h). These interval ranges seem very large, but when we looked at participants' performance in detail, we found that the large majority of the data points were below 20 cm (see Fig. 3), with just 5.6% of the values being above this threshold. In fact, the 5%-trimmed means for PRECISION-OFFSET were 7.1 cm and 6.3 cm, respectively, for the two TIME-OF-RECALL conditions. A Repeated-Measures ANOVA showed a statistically significant effect of TIME-OF-RECALL on PRECISION-OFFSET ($F_{(1,16)} = 7.832$, $p = .013 < .05$), showing that participants were more precise on the second day, yet only slightly (by 6 mm). The same test found no significant effect of CHANNEL on PRECISION-OFFSET ($F_{(1.445,23.126)} = 1.549$, $p = .088 > .05$, *n.s.*, Greenhouse–Geisser $\hat{\epsilon} = .181$). These findings show that the upper limit of human precision to reach mid-air loci with no feedback other than memory alone can be reasonably approximated at

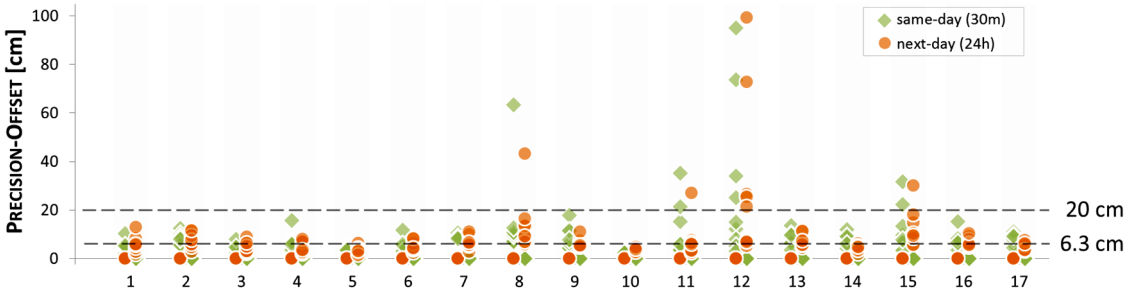

Fig. 3. Precision offset values, in centimeters, shown for all the trials performed by all participants on the *same-day* and *next-day*. The average performance is 6.35 cm ($CI_{95\%}$ = [5.99, 6.72] cm), while 95% of all the data fall below the 20 cm threshold.

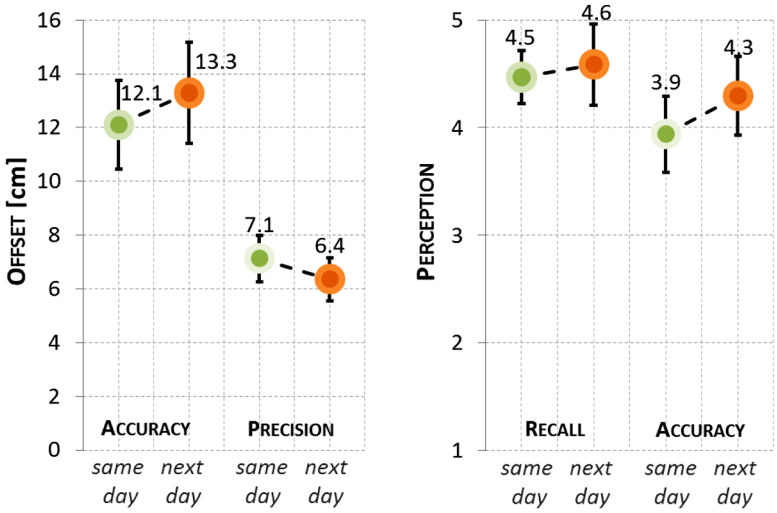

Fig. 4. Mean accuracy and precision offsets (left) and participants' perceived recall and accuracy (right) for each TIME-OF-RECALL experimental condition. Error bars show 95% CIs.

20 cm with the average performance staying around 6.35 cm ($CI_{95\%}$ = [5.99, 6.72] cm). A Kruskal–Wallis test found no significant effect of participants' spatial orientation skills on PRECISION-OFFSET ($\chi^2(2) = 1.348$, $p = .510 > .05$, *n.s.* for the *same-day* condition and $\chi^2(2) = 1.348$, $p = .510 > .05$, *n.s.* for *next-day*, respectively).

We continue our analysis with the ACCURACY-OFFSET to understand how much participants were off in their pointing with respect to the actual channels' loci. By using the 20 cm upper threshold informed by the previous analysis, we were able to estimate recall rates at 80.4% and 79.1%, respectively, for the two TIME-OF-RECALL conditions. (Any trial for which the distance between the pointed locus and the original one, as set during the configuration phase, was larger than 20 cm was considered a misrecall.) For the valid trials, ACCURACY-OFFSET varied between 2.2 cm and 19.0 cm ($M = 8.9$, $SD = 3.4$) during the *same-day* and between 1.9 cm and 19.4 cm ($M = 9.8$, $SD = 4.7$) on the *next-day*. A Repeated Measures ANOVA test found no significant effect of TIME-OF-RECALL ($F_{(1,16)} = 3.179$, $p = .094 > .05$, *n.s.*) or CHANNEL ($F_{(4.138,66.214)} = 0.748$, $p = .567 > .05$, *n.s.*, Greenhouse-Geisser $\hat{\epsilon} = .517$) on ACCURACY-OFFSET. A Kruskal-Wallis test found no significant effect of participants' spatial skills on ACCURACY-OFFSET ($\chi^2(2) = 0.182$, $p = .913 > .05$, *n.s.* for the *same-day* condition and $\chi^2(2) = 0.058$, $p = .972 > .05$, *n.s.* for *next-day*, respectively).

5.3. Perceived recall and accuracy performance

Participants rated their performance (recall and accuracy) after each recall phase using 5-point Likert scales. We found no significant effect of TIME-OF-RECALL on PERCEIVED-RECALL (median values 4 and 5, mean values 4.5 and 4.6, $Z_{(N=17)} = -0.632$, $p = .527 > .05$, *n.s.*). However, we did find a statistically significant effect of TIME-OF-RECALL on PERCEIVED-ACCURACY ($Z_{(N=17)} = -2.449$, $p = .014 < .05$, $r = .420$). Fig. 4, right shows the mean values for participants' self-perception ratings.[3] Both ratings for the ability to recall and to reach loci accuracy were slightly larger for the *next-day* phase than for the *same-day* one. This result correlates with the analysis of objective offsets, where we observed that participants were in fact slightly more precise on the second day.

5.4. Detecting pointing actions in mid-air

We wanted to know how accurate Hover can detect users' reaching and pointing actions to channels floating in mid-air. To this end, we computed recognition rates by comparing participants' trials on the *same-day* and from *next-day* with the original spatial loci they had set during the configuration stage. Our procedure implemented a Nearest-Neighbor classification that assigned the locus of a trial $p = (x, y, z)$ to the closest channel in space in terms of the Euclidean distance:

$$\text{Assign } p \text{ to channel } k \text{ if } ||p - \mu_k|| = \min_{i=1..9} ||p - \mu_i|| \quad (9)$$

where μ_i goes through all the channels. For this analysis we didn't

[3] Although medians are the appropriate statistics for reporting Likert-scale ratings, we decided to report both median and mean values to better highlight the difference between the two conditions. For example, note how median PERCEIVED-ACCURACY values are 4 for both the *same-day* and *next-day* conditions, yet the difference between them is statistically significant. Mean values indicate the direction of the difference.

consider the trials for which participants misrecalled the loci where they had originally positioned channels in mid-air, *i.e.*, trials for which $\min_i ||p - \mu_i|| > 20$ cm; see our previous results and discussion about participants' performance evaluated with the PRECISION and ACCURACY offset measures.

Obviously, the layout type impacts recognition rates. For instance, layouts with channels that are separated by large distances will clearly result in higher recognition rates, little influenced by the precision of the user to reach specific loci in space. Therefore, in order to reduce the influence of the layout (on which we had no control during the experiment), we analyzed all participants' proposed layouts to see how many respect the 20 cm precision threshold informed by our previous analysis. Let n_e denote the number of "effective" channels of a layout for which MIN-PROXIMITY is at least 20 cm. We started with $n_e = 9$ and applied the following steps: remove each channel from the layout and, if the 20 cm condition is met for n_e, keep the layout with n_e channels that maximizes AVG-PROXIMITY; otherwise, decrement n_e. The number of effective channels for each participant's layout is illustrated in Table 1 next to the numerical characterization of the layout using proximity, volume, and spread measurements. On average, the number of effective channels was 6.9 ($SD = 2.2$), representing a 23.3% reduction from the original number. Interestingly, according to a Smartclip, LG, and Nielsen study (Smartclip, 2015) (August 2015, UK respondents), the average number of TV channels that are actually followed is 7.9 for smart TV owners and 7.1 for traditional TV owners, respectively.

Classification results of participants' pointing trials to mid-air loci showed 97.6% recognition accuracy on the *same-day* and 99.4% on the *next-day*, respectively; see Table 1 for individual results for each participant. To understand recognition performance for the maximum number of channels considered originally (9), we also computed recognition rates for the six participants that proposed layouts with 9 effective channels (i.e., P_{12}, P_4, P_1, P_{16}, P_{15}, and P_7; see Table 1). Using this data, we found that Hover successfully detected users' pointing locations in mid-air with nearly perfect accuracy, 98.9% on the *same-day* and 100% on the *next-day*, respectively.

6. Discussion and future work

We conduct in this section a discussion of Hover and our experimental results in terms of the interaction qualities desirable for user interfaces based on air pointing, such as learnability, expressivity, and comfort (Cockburn et al., 2011). We also present suggestions for memorization techniques to support recall (Perrault et al., 2015; Worthen and Hunt, 2017) and, therefore, increase users' performance with Hover. We also suggest ideas for future work extensions regarding strategies to deal with false positives during the recognition of air pointing actions for practical systems implementing Hover.

6.1. The interaction qualities of Hover

Hover relies on users' abilities to recall associations between their preferred TV channels and loci from the room and point to those loci accurately. From this perspective, Hover falls into the category of "air pointing" interaction techniques, which were examined by Cockburn et al. (2011) in terms of interaction dimensions and qualities. Specifically, Cockburn et al. (2011) identified five interaction dimensions for air pointing, *i.e., reference frame, scale of spatial input, input degrees of freedom, feedback modality*, and *feedback content* (p. 405). It is interesting therefore to see how Hover positions with respect to these dimensions. In terms of the *reference frame* for spatial input, our participants were given absolute freedom to choose spatial locations to anchor their preferred TV channels in the room. Locations actually chosen by participants included absolute locations, relative to the world (*e.g.*, P_{10} pointed to books from a shelf found in the room), relative to the body (*e.g.*, P_9 placed all the channels on their body: on shoulders, head, waist, hands, and knees; see Fig. 2), and relative to devices (*e.g.*, P_3 used the four corners of the TV frame as visual targets to guide their pointing actions). In terms of the *scale dimension*, air pointing movements performed by our participants were large when referring to objects from the extrapersonal space and had smaller amplitude when pointing to mid-air targets in front or near the body in the peripersonal and personal space. The number of *degrees of freedom* were three for our experimental setup and implementation using the Vicon motion capture system that reported the *x, y*, and *z* coordinates of the pointing finger, which was sufficient for the purpose of our experiment to understand our participants' accuracy and recall performance. However, practical implementations of Hover may require more degrees of freedom, such as tracking the entire arm (*i.e.*, multiple points on the arm, not just the index finger) or may use different kind of data, such as accelerations and orientations provided by a smart armband (Haque et al., 2015; Popovici and Vatavu, 2018) worn on the forearm or by a smartwatch worn on the wrist (Katsuragawa et al., 2016); such engineering aspects are left for examination in future work, according to the context of use in which Hover is to be applied.

Applied Hover interfaces will require design regarding *feedback modality* and *feedback content*, two interaction dimensions listed by Cockburn et al. (2011) for air pointing that we did not require and, therefore, did not implement in our experiment. However, a deployment of Hover in a practical scenario will provide feedback to viewers following their air pointing actions, from the channel changing as minimal feedback to suggestions and opportunities for corrective actions when the outcome of the air pointing command was other than intended. Regarding the latter, false positives (*i.e.*, gesture commands identified accidentally by the system and not intended by users) need to be dealt with accordingly, as they can affect negatively the user experience by creating the perception of a system out of the user's control. This problem can be addressed in various ways during design, *e.g.*, MAGIC 2.0 (Kohlsdorf et al., 2011) is a technique for false positive prediction and prevention that informs practitioners, right from the design process, about the possibility of a candidate gesture to trigger accidentally during use. Another approach is to design gesture movements that are low-false positives in the first place (Kawahata et al., 2016). However, these techniques do not apply when users customize the gesture set to their own preferences, in which case a preliminary analysis of the gesture set or the air pointing movements proposed by users could be performed to detect likely false positives or misrecognitions (*i.e.*, a post-hoc MAGIC 2.0 test) and, if needed, advise users to reconsider their preferences. The problem of false positives and misrecognition becomes even more important for social television watching, when multiple persons are present in the room, watching the same TV show, but also talking to each other and making use of gestures and body movements during their conversations, directed either to the interlocutors or to the TV. In fact, gesture user interfaces are flexible enough to enable multiple users to share control of the TV set (Vatavu and Pentiuc, 2008). However, it is also a known fact that gestures articulated by different viewers may turn out contradicting each other, resulting in ambiguity and an unknown state for the system (Plaumann et al., 2016). Consequently, solving possible conflicts in a multi-user gesture-based TV control environment requires personalized mediation strategies to be adopted, such as the "master user" or "varying master users" scenarios, implemented by the TV user interface; see Plaumann et al. (2016). A feature to turn on and off gesture recognition may also be useful to handle false positives. We leave such practical aspects for future exploration.

We also use the six interaction qualities for air pointing techniques proposed by Cockburn et al. (2011), *e.g., learnability for novices, selection speed for experts, accuracy, expressivity, cognitive effort*, and *comfort* (p. 404), to analyze Hover. First, Hover permits users to configure their own associations between spacial locations and TV channels and, thus, *learnability for novices* is implicitly assured. Although we do not have supporting data, we expect based on prior work (Perrault et al., 2015) that users become more accurate and faster with mid-air pointing to

spatial locations with practice and time, which is in accord with the *selection speed for experts* and *accuracy* interaction qualities. Hover was tested with nine spatial locations, which we considered to be a sufficient number of shortcuts to preferred TV channels, an expectation confirmed by our participants' feedback regarding the number of channels they actually followed on a regular basis, *i.e.*, 5.1 channels on average; see the "Favorite channels" column of Table 1. From this perspective, Hover assures sufficient *expressivity* for users, although future work is recommended to understand the upper limits for Hover in terms of the number of spatial locations that can be effectively operated. Although we did not tap into the *cognitive effort* and *comfort* interaction qualities explicitly, our PERCEIVED-RECALL and PERCEIVED-ACCURACY measures indicated positive results, *i.e.*, 4.5 mean rating for recall (on a scale from "1, very difficult" to "5, very easy") and 4.1 for perceived accuracy (on a scale from "1, not accurate at all" to "5, very accurate"). Based on our data and the above analysis, we conclude that Hover satisfies reasonably well the six interaction qualities of Cockburn et al. (2011) desirable for user interfaces based on air pointing.

6.2. Memorization techniques and extending Hover to more targets

With Hover, targets exist entirely in the user's memory and, consequently, user performance is determined by the ability to learn and recall mid-air loci as accurately as possible. Memory is a brain-wide distributed domain-specific process made possible by different regions of the brain working in conjunction to provide us the ability to encode, store, and recall information at will (Squire and Wixted, 2011). Recall of a particular piece of memory depends on how well the same neocortical regions that generated the encoding in the first place can be effectively reactivated on demand (Squire and Wixted, 2011). According to Miller's observations, the capacity of the short-term memory is limited to handling between 5 and 9 independent items, *i.e.*, the magical number 7 ± 2 (Miller, 1956). Even for learned material, the decline of memory bears an exponential dependence with time according to Ebbinghaus' "forgetting curve hypothesis:" with no conscious review of the learned material, humans tend to nearly halve their memory of newly learned knowledge in a few days (Karpicke and Blunt, 2011; Murre and Dros, 2015). Consequently, memorization techniques are needed when operating with a large number of loci, such as those discussed by Worthen and Hunt (2017) and Perrault et al. (2015).

As we found in our experiment, users exhibit various levels of performance in terms of their precision, recall, and spatial abilities and, also, they have various preferences for layouts of channels anchored in mid-air. This result directly impacts the complexity of memory-based spatial user interfaces, such as Hover, which should be adapted to each user. In particular, the actual number of active loci needs specific consideration. We conducted our experimental evaluation of Hover with a number of nine loci, for reasons that we outlined before, but our participants indicated during the experiment that they followed less than nine channels ($M = 5.1$, $SD = 2.0$). The Nielsen Total Audience Report (2016) reported that the percentage of television channels actually viewed is 9.6% of all receivable channels, which represents 20 channels at maximum out of a selection of say 200 offered by the cable provider (Nielsen, 2016). This figure is viable for the human memory, as prior work demonstrated nearly perfect recall of 48 physical loci after one week using memorization techniques (Perrault et al., 2015). Also, they show that the number of useful shortcuts to TV channels may actually be limited in practice, which makes Hover-like systems practical for users to learn and recall just a few spatial locations. Besides shortcuts to preferred channels, mid-air loci could be assigned to other smart TV functions, such as opening the web browser, starting applications, shifting between various picture mode settings, such as movie mode, and so on; see Vatavu (2012b) and Vatavu and Zaiţi (2014) for a wide range of smart TV functions for which gesture commands were examined.

7. Conclusion

We introduced Hover, a concept and interface for the remote control of the TV set that operates based on air pointing in the users' personal, peripersonal, and extrapersonal space with the support of cognitive maps formed, stored into, and retrieved from users' memory regarding associations between spacial locations and TV channels. Our evaluation revealed the human performance for recalling and reaching to loci in mid-air, which we examined to understand practical scenarios where Hover can be implemented effectively, *i.e.*, with 99% recognition rate for air pointing movements. With Hover, we hope to bring a new perspective in the interactive TV community, which will inspire practitioners to consider spatial locations and mid-air for the design of new interactive experiences for our future smart home entertainment environments centered on the smart TV.

Acknowledgments

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI - UEFISCDI, project number PN-III-P3-3.1-PM-RO-CN-2018-0032 (3BM/2018), within PNCDI III. Research was conducted in the Machine Intelligence and Information Visualization Lab (MintViz) of the MANSiD Research Center. The infrastructure was provided by the University of Suceava and was partially supported from the project *"Integrated center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication and control"*, no. 671/09.04.2015, Sectoral Operational Program for Increase of the Economic Competitiveness, co-funded from the European Regional Development Fund.

References

Ali-Hasan, N., Soto, B., 2015. 8 things to consider when designing interactive TV experiences. TVX 2015, the ACM International Conference on Interactive Experiences for Television and Online Video. https://ai.google/research/pubs/pub43865.

Bailly, G., Vo, D.-B., Lecolinet, E., Guiard, Y., 2011. Gesture-aware remote controls: guidelines and interaction technique. Proceedings of the 13th International Conference on Multimodal Interfaces. ACM, New York, NY, USA, pp. 263–270. https://doi.org/10.1145/2070481.2070530.

Bergstrom-Lehtovirta, J., Boring, S., Hornbæk, K., 2017. Placing and recalling virtual items on the skin. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1497–1507. https://doi.org/10.1145/3025453.3026030.

Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M., 2008. Trends in the living room and beyond: results from ethnographic studies using creative and playful probing. Comput. Entertain. 6 (1), 5:1–5:23. https://doi.org/10.1145/1350843.1350848.

Bernhaupt, R., Pirker, M.M., 2014. User interface guidelines for the control of interactive television systems via smart phone applications. Behav. Inf. Technol. 33 (8), 784–799. https://doi.org/10.1080/0144929X.2013.810782.

Bobeth, J., Schrammel, J., Deutsch, S., Klein, M., Drobics, M., Hochleitner, C., Tscheligi, M., 2014. Tablet, gestures, remote control?: influence of age on performance and user experience with iTV applications. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 139–146. https://doi.org/10.1145/2602299.2602315.

Bolt, R.A., 1980. Put-that-there: voice and gesture at the graphics interface. SIGGRAPH Comput. Graph. 14 (3), 262–270. https://doi.org/10.1145/965105.807503.

Brown, A., Jones, R., Crabb, M., Sandford, J., Brooks, M., Armstrong, M., Jay, C., 2015. Dynamic subtitles: the user experience. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 103–112. https://doi.org/10.1145/2745197.2745204.

Chagas, D.A., Furtado, E.S., 2013. MoveRC: Attention-aware remote control. Proceedings of the 19th Brazilian Symposium on Multimedia and the Web. ACM, New York, NY, USA, pp. 277–280. https://doi.org/10.1145/2526188.2526235.

Choi, S., Han, J., Lee, G., Lee, N., Lee, W., 2011. RemoteTouch: touch-screen-like interaction in the TV viewing environment. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 393–402. https://doi.org/10.1145/1978942.1978999.

Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., Looser, J., 2011. Air pointing: design and evaluation of spatial target acquisition with and without visual feedback. Int. J. Hum.-Comput. Stud. 69 (6), 401–414. https://doi.org/10.1016/j.ijhcs.2011.02.005.

Cognitrom, 2018. Cognitrom Assessment System, CAS++. http://www.cognitrom.ro/produs/evaluare-psihologica/.

Colley, A., Rantakari, J., Häkkilä, J., 2014. Augmenting the home to remember: initial

user perceptions. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, New York, NY, USA, pp. 1369–1372. https://doi.org/10.1145/2638728.2641717.
Cox, D., Wolford, J., Jensen, C., Beardsley, D., 2012. An evaluation of game controllers and tablets as controllers for interactive TV applications. Proceedings of the 14th ACM International Conference on Multimodal Interaction. ACM, New York, NY, USA, pp. 181–188. https://doi.org/10.1145/2388676.2388713.
Dezfuli, N., Khalilbeigi, M., Huber, J., Müller, F., Mühlhäuser, M., 2012. PalmRC: imaginary palm-based remote control for eyes-free television interaction. Proceedings of the 10th European Conference on Interactive TV and Video. ACM, New York, NY, USA, pp. 27–34. https://doi.org/10.1145/2325616.2325623.
Dim, N.K., Silpasuwanchai, C., Sarcar, S., Ren, X., 2016. Designing mid-air TV gestures for blind people using user- and choice-based elicitation approaches. Proceedings of the 2016 ACM Conference on Designing Interactive Systems. ACM, New York, NY, USA, pp. 204–214. https://doi.org/10.1145/2901790.2901834.
Fruchard, B., Lecolinet, E., Chapuis, O., 2018. Impact of semantic aids on command memorization for on-body interaction and directional gestures. Proceedings of the 2018 International Conference on Advanced Visual Interfaces. ACM, New York, NY, USA, pp. 14:1–14:9. https://doi.org/10.1145/3206505.3206524.
Geerts, D., Leenheer, R., De Grooff, D., Negenman, J., Heijstraten, S., 2014. In front of and behind the second screen: viewer and producer perspectives on a companion app. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 95–102. https://doi.org/10.1145/2602299.2602312.
Gheran, B.-F., Vanderdonckt, J., Vatavu, R.-D., 2018. Gestures for smart rings: empirical results, insights, and design implications. Proceedings of the 2018 Designing Interactive Systems Conference. ACM, New York, NY, USA, pp. 623–635. https://doi.org/10.1145/3196709.3196741.
Guerreiro, T., Gamboa, R., Jorge, J., 2008. Mnemonical body shortcuts: improving mobile interaction. Proceedings of the 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction. ACM, New York, NY, USA, pp. 11:1–11:8. https://doi.org/10.1145/1473018.1473033.
Gunther, R., 2011. Rethinking the television experience. UX Magazine, https://uxmag.com/articles/rethinking-the-television-experience.
Gustafson, S., Bierwirth, D., Baudisch, P., 2010. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 3–12. https://doi.org/10.1145/1866029.1866033.
Gustafson, S., Holz, C., Baudisch, P., 2011. Imaginary phone: learning imaginary interfaces by transferring spatial memory from a familiar device. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 283–292. https://doi.org/10.1145/2047196.2047233.
Gustafson, S.G., Rabe, B., Baudisch, P.M., 2013. Understanding palm-based imaginary interfaces: the role of visual and tactile cues when browsing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 889–898. https://doi.org/10.1145/2470654.2466114.
Haque, F., Nancel, M., Vogel, D., 2015. Myopoint: pointing and clicking using forearm mounted electromyography and inertial motion sensors. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 3653–3656. https://doi.org/10.1145/2702123.2702133.
Harrison, C., Faste, H., 2014. Implications of location and touch for on-body projected interfaces. Proceedings of the 2014 Conference on Designing Interactive Systems. ACM, New York, NY, USA, pp. 543–552. https://doi.org/10.1145/2598510.2598587.
Hartmann, J., Vogel, D., 2018. An evaluation of mobile phone pointing in spatial augmented reality. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. LBW122:1–LBW122:6. https://doi.org/10.1145/3170427.3188535.
Heijboer, M., van den Hoven, E., Bongers, B., Bakker, S., 2016. Facilitating peripheral interaction: design and evaluation of peripheral interaction for a gesture-based lighting control with multimodal feedback. Pers. Ubiquitous Comput. 20 (1), 1–22. https://doi.org/10.1007/s00779-015-0893-5.
Hincapié-Ramos, J.D., Guo, X., Irani, P., 2014. The consumed endurance workbench: a tool to assess arm fatigue during mid-air interactions. Proceedings of the 2014 Companion Publication on Designing Interactive Systems. ACM, New York, NY, USA, pp. 109–112. https://doi.org/10.1145/2598784.2602795.
Holmes, N.P., Spence, C., 2004. The body schema and the multisensory representation(s) of peripersonal space. Cogn Process 5 (2), 94–105. https://dx.doi.org/10.1007%2Fs10339-004-0013-3
Holz, C., Bentley, F., Church, K., Patel, M., 2015. "I'm just on my phone and they're watching TV": quantifying mobile device use while watching television. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 93–102. https://doi.org/10.1145/2745197.2745210.
Jang, S., Stuerzlinger, W., Ambike, S., Ramani, K., 2017. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 3328–3339. https://doi.org/10.1145/3025453.3025523.
Jeong, S., Song, T., Kwon, K., Jeon, J.W., 2012. TV remote control using human hand motion based on optical flow system. Proceedings of the 12th International Conference on Computational Science and Its Applications - Volume Part III. Springer-Verlag, Berlin, Heidelberg, pp. 311–323. https://doi.org/10.1007/978-3-642-31137-6_24.
Jiang, J., Jeng, W., He, D., 2013. How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search. Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, NY, USA, pp. 143–152. https://doi.org/10.1145/2484028.2484092.
Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., Raghuvanshi, N., Shapira, L., 2014. RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 637–644. https://doi.org/10.1145/2642918.2647383.
Jones, B.R., Benko, H., Ofek, E., Wilson, A.D., 2013. IllumiRoom: peripheral projected illusions for interactive experiences. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 869–878. https://doi.org/10.1145/2470654.2466112.
Karpicke, J.D., Blunt, J.R., 2011. Retrieval practice produces more learning than elaborative studying with concept mapping. 331 (6018), 772–775. https://doi.org/10.1126/science.1199327.
Katsuragawa, K., Pietroszek, K., Wallace, J.R., Lank, E., 2016. Watchpoint: freehand pointing with a smartwatch in a ubiquitous display environment. Proceedings of the International Working Conference on Advanced Visual Interfaces. ACM, New York, NY, USA, pp. 128–135. https://doi.org/10.1145/2909132.2909263.
Kawahata, R., Shimada, A., Yamashita, T., Uchiyama, H., Taniguchi, R.-I., 2016. Design of a low-false-positive gesture for a wearable device. Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods. SciTePress, pp. 581–588.
Kohlsdorf, D., Starner, T., Ashbrook, D., 2011. MAGIC 2.0: a web tool for false positive prediction and prevention for gesture recognition systems. Face and Gesture 2011. pp. 1–6. https://doi.org/10.1109/FG.2011.5771412.
LeapMotion, 2018. Leap Motion. https://www.leapmotion.com/.
Lee, D.-H., Kim, J.-H., Kim, H.-Y., Park, D.-Y., 2016. Remote application control technology and implementation of HTML5-based Smart TV platform. Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi Media. ACM, New York, NY, USA, pp. 208–211. https://doi.org/10.1145/3007120.3007159.
LG, 2018. LG Magic Remote: TV remotes for Smart TVs. http://www.lg.com/us/magic-remote.
Li, F.C.Y., Dearman, D., Truong, K.N., 2009. Virtual Shelves: interactions with orientation aware devices. Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 125–128. https://doi.org/10.1145/1622176.1622200.
Lin, S.-Y., Shie, C.-K., Chen, S.-C., Hung, Y.-P., 2013. AirTouch Panel: are-anchorable virtual touch panel. Proceedings of the 21st ACM International Conference on Multimedia. ACM, New York, NY, USA, pp. 625–628. https://doi.org/10.1145/2502081.2502164.
Lindlbauer, D., Grønbæk, J.E., Birk, M., Halskov, K., Alexa, M., Müller, J., 2016. Combining shape-changing interfaces and spatial augmented reality enables extended object appearance. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 791–802. https://doi.org/10.1145/2858036.2858457.
Lu, S.-P., Florea, R.-M., Cesar, P., Schelkens, P., Munteanu, A., 2017. Efficient depth-aware image deformation adaptation for curved screen displays. Proceedings of the on Thematic Workshops of ACM Multimedia 2017. ACM, New York, NY, USA, pp. 442–450. https://doi.org/10.1145/3126686.3126692.
McGill, M., Williamson, J.H., Brewster, S., 2016. Examining the role of smart TVs and VR HMDs in synchronous at-a-distance media consumption. ACM Trans. Comput.-Hum. Interact. 23 (5), 33:1–33:57. https://doi.org/10.1145/2983530.
Mehrotra, S., 2018. Potmote: a TV remote control for older adults. Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, New York, NY, USA, pp. 486–488. https://doi.org/10.1145/3234695.3240989.
Microsoft, 2018. Kinect - Windows app development. https://developer.microsoft.com/en-us/windows/kinect.
Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2), 81.
Montano Murillo, R.A., Subramanian, S., Martinez Plasencia, D., 2017. Erg-O: Ergonomic optimization of immersive virtual environments. Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 759–771. https://doi.org/10.1145/3126594.3126605.
Murnane, K., 2018. Google's assistant has found a new home in LG's OLED and Super UHD TVs. https://www.forbes.com/sites/kevinmurnane/2018/01/03/googles-assistant-has-found-a-new-home-in-lgs-oled-and-super-uhd-tvs.
Murre, J.M., Dros, J., 2015. Replication and analysis of Ebbinghaus' forgetting curve. PLoS One 10 (7). https://doi.org/10.1371/journal.pone.0120644.
Neate, T., Evans, M., Jones, M., 2017. Enhancing interaction with dual-screen television through display commonalities. Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 91–103. https://doi.org/10.1145/3077548.3077549.
Nielsen, 2016. The Nielsen total audience report: Q2 2016. http://www.nielsen.com/us/en/insights/reports/2016/the-nielsen-total-audience-report-q2-2016.html.
Perrault, S.T., Lecolinet, E., Bourse, Y.P., Zhao, S., Guiard, Y., 2015. Physical Loci: leveraging spatial, object and semantic memory for command selection. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 299–308. https://doi.org/10.1145/2702123.2702126.
Plaumann, K., Lehr, D., Rukzio, E., 2016. Who has the force?: solving conflicts for multi user mid-air gestures for TVs. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 25–29. https://doi.org/10.1145/2932206.2932208.
Popovici, I., Vatavu, R.-D., 2018. Perceived usability, desirability, and workload of mid-air gesture control for smart TVs. Proceedings of the 15th International Conference on

Human Computer Interaction. MatrixRom, Bucharest, pp. 91–98. https://dblp.org/rec/conf/rochi/PopoviciV18

Rao, J., Ture, F., He, H., Jojic, O., Lin, J., 2017. Talking to your TV: context-aware voice search with hierarchical recurrent neural networks. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, New York, NY, USA, pp. 557–566. https://doi.org/10.1145/3132847.3132893.

Rateau, H., Grisoni, L., De Araujo, B., 2014. Mimetic interaction spaces: controlling distant displays in pervasive environments. Proceedings of the 19th International Conference on Intelligent User Interfaces. ACM, New York, NY, USA, pp. 89–94. https://doi.org/10.1145/2557500.2557545.

Renger, B., Feng, J., Dan, O., Chang, H., Barbosa, L., 2011. VoiSTV: voice-enabled social TV. Proceedings of the 20th International Conference Companion on World Wide Web. ACM, New York, NY, USA, pp. 253–256. https://doi.org/10.1145/1963192.1963302.

Rutledge, R.L., Massey, A.K., Antón, A.I., 2016. Privacy impacts of IoT devices: a SmartTV case study. Proceedings of the 24th IEEE International Requirements Engineering Conference Workshops. pp. 261–270. https://doi.org/10.1109/REW.2016.050.

Samsung, 2018. Samsung Smart TV: TV gesture book. http://www.samsung.com/ph/smarttv/common/guide_book_3p_si/waving.html.

Schipor, O.-A., Vatavu, R.-D., 2018. Invisible, inaudible, and impalpable: users' preferences and memory performance for digital content in thin air. IEEE Pervasive Comput. 17 (4), 76–85. https://doi.org/10.1109/MPRV.2018.2873856.

Shoemaker, G., Tsukitani, T., Kitamura, Y., Booth, K.S., 2010. Body-centric interaction techniques for very large wall displays. Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. ACM, New York, NY, USA, pp. 463–472. https://doi.org/10.1145/1868914.1868967.

Smartclip, 2015. Smart TV insights 2015: a study about the Smart TV market and the changing TV consumption. https://tinyurl.com/ydxmvwun.

Squire, L.R., Wixted, J.T., 2011. The cognitive neuroscience of human memory since H.M. Annu. Rev. Neurosci. 34, 259–288. https://doi.org/10.1146/annurev-neuro-061010-113720.

Steins, C., Gustafson, S., Holz, C., Baudisch, P., 2013. Imaginary devices: gesture-based interaction mimicking traditional input devices. Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 123–126. https://doi.org/10.1145/2493190.2493208.

Tokuda, Y., Norasikin, M.A., Subramanian, S., Martinez Plasencia, D., 2017. MistForm: adaptive shape changing fog screens. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 4383–4395. https://doi.org/10.1145/3025453.3025608.

Tolman, E., 1948. Cognitive maps in rats and men. Psychol. Rev. 55 (4), 189–208. https://doi.org/10.1037/h0061626.

Toyoura, M., Shono, T., Mao, X., 2010. BioMetal Glove. Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology. ACM, New York, NY, USA, pp. 135–138. https://doi.org/10.1145/1889863.1889892.

Vatavu, R.-D., 2012. Point & click mediated interactions for large home entertainment displays. Multimed. Tools Appl. 59 (1), 113–128. https://doi.org/10.1007/s11042-010-0698-5.

Vatavu, R.-D., 2012. User-defined gestures for free-hand TV control. Proceedings of the 10th European Conference on Interactive TV and Video. ACM, New York, NY, USA, pp. 45–48. https://doi.org/10.1145/2325616.2325626.

Vatavu, R.-D., 2013. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. J. Ambient Intell. Smart Environ. 5 (2), 187–211. https://doi.org/10.3233/AIS-130200.

Vatavu, R.-D., 2013. There's a world outside your TV: exploring interactions beyond the physical TV screen. Proceedings of the 11th European Conference on Interactive TV and Video. ACM, New York, NY, USA, pp. 143–152. https://doi.org/10.1145/2465958.2465972.

Vatavu, R.-D., 2015. Audience silhouettes:peripheral awareness of synchronous audience kinesics for social television. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 13–22. https://doi.org/10.1145/2745197.2745207.

Vatavu, R.-D., 2017. Smart-Pockets: body-deictic gestures for fast access to personal data during ambient interactions. Int. J. Hum.-Comput. Stud. 103 (C), 1–21. https://doi.org/10.1016/j.ijhcs.2017.01.005.

Vatavu, R.-D., Mancas, M., 2014. Visual attention measures for multi-screen TV. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 111–118. https://doi.org/10.1145/2602299.2602305.

Vatavu, R.-D., Mossel, A., Schönauer, C., 2016. Digital vibrons: understanding users' perceptions of interacting with invisible, zero-weight matter. Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 217–226. https://doi.org/10.1145/2935334.2935364.

Vatavu, R.-D., Pentiuc, S.-G., 2008. Interactive coffee tables: interfacing TV within an intuitive, fun and shared experience. Proceedings of the 6th European Conference on Changing Television Environments. Springer-Verlag, Berlin, Heidelberg, pp. 183–187. https://doi.org/10.1007/978-3-540-69478-6_24.

Vatavu, R.-D., Zaiţi, I.-A., 2014. Leap gestures for TV:insights from an elicitation study. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. ACM, New York, NY, USA, pp. 131–138. https://doi.org/10.1145/2602299.2602316.

Vogel, D., Balakrishnan, R., 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 137–146. https://doi.org/10.1145/1029632.1029656.

Worthen, J.B., Hunt, R.R., 2017. Mnemonic Techniques: Underlying Processes and Practical Applications. In: Byrne, J.H. (Ed.), Learning and Memory: A Comprehensive Reference (2nd Edition). Academic Press, Oxford, pp. 515–527. https://doi.org/10.1016/B978-0-12-809324-5.21063-8.

Yan, Y., Yu, C., Ma, X., Huang, S., Iqbal, H., Shi, Y., 2018. Eyes-free target acquisition in interaction space around the body for virtual reality. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 42:1–42:13. https://doi.org/10.1145/3173574.3173616.

Yusufov, M., Kornilov, I., 2013. Roles of Smart TV in IoT-environments: a survey. Proceedings of the 13th Conference of Open Innovations Association. pp. 163–168. https://doi.org/10.23919/FRUCT.2013.8124240.

Zaiţi, I.-A., Pentiuc, Ş.-G., Vatavu, R.-D., 2015. On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. Pers. Ubiquitous Comput. 19 (5), 821–838. https://doi.org/10.1007/s00779-015-0863-y.