The Impact of Comfortable Viewing Positions on Smart TV Gestures

Jeroen Vanattenhoven*, David Geerts†, Jean Vanderdonckt‡ and Jorge-Luis Pérez-Medina§
**KU Leuven*, Oude Markt 13, B-3000 Leuven, Belgium
†Meaningful Interactions Lab (Mintlab), *KU Leuven*, Parkstraat, 45 Bus 3605, B-3000 Leuven, Belgium
‡Louvain Research Inst. in Management and Org., *Université catholique de Louvain*, B-1348 Louvain-la-Neuve, Belgium
§Intelligent and Interactive Systems Laboratory (SI2-Lab), *Universidad de Las Américas*, Quito, Ecuador

***Abstract*—Whereas gesture elicitation studies for TV interaction assume that participants adopt an upright, frontal viewing position, we asked 21 participants to hold a natural, comfortable viewing position, the posture they adopt when watching TV at home. By involving a broad selection of users regarding age, profession, our study targets a higher ecological validity than existing studies. Agreements rates were lower than existing studies using an upright, frontal viewing position. Participants experienced problems due to (1) having to use their slave hand instead of their dominant hand, (2) being in a certain orientation with their head making it more difficult to perform some physical movements, and (3) being hindered in their movement by the sofa there lay on. Since each person may have a different position inducing different gestures due to the aforementioned problems, the effect of comfortable viewing position is analyzed by comparison to gestures for a frontal position.**

***Index Terms*—Interactive TV, Gesture elicitation study, Gesture user interface, Natural interaction.**

I. Introduction

People have been using the remote control to operate their television sets for over 30 years, and they still do. Currently, some alternatives exist but none has been able to take over the dominant role of the remote control yet. Notwithstanding this dominance, many problems subsist with remote controls [1]: they can get lost in the home, they require batteries, and they are confused or interchanged with the numerous other remote controls for set top boxes, gaming consoles, and other ambient devices. Two main alternatives are voice and gesture interaction, for which the research community and the industry are very active: the body of literature on gesture interaction has grown substantially and several products are available that allow people to control their TV set vocally or gesturally, such as with hand gestures [2]. Gesture interaction does not require batteries, nor do you need to manage several devices. but can result in fatigue with time. Researchers are still investigating the best mapping between an action to be executed by the system, which is called a "referent" [3], [4] and the corresponding gesture. Referent-gesture mapping is often devised by the system maker. Instead, methods have been developed to involve people in establishing this mapping. In gesture elicitation studies [3], [5], representative users are shown the effect of a gesture, and asked to perform its cause. Repeating this exercise with many participants results in a large number of gestures for each referent. Several user and gesture elicitation studies are available for TV interaction. When reviewing their findings and methodologies, we noticed that, to the best of our knowledge, participants were always asked to assume an upright, frontal viewing position before the TV. In order to increase the ecological validity [6], we opted to repeat such a study while instructing participants to choose their viewing position as part of the elicitation process instead of imposing it. As watching television is still very much a leisure activity, people will be more likely to assume a more comfortable, natural viewing position, compared to the upright, frontal viewing position. These viewing positions are likely to restrict the movement of the body parts used to perform the gestures. Furthermore, people might not always be able to use their dominant hand or arm due to their specific position. Therefore, we believe that the difference between the lab and the real context of use might be quite significant.

For this purpose, this paper is organized as follows: Section II reports on prior work done in the field of gesture elicitation studies in general, with a focus on smart TV; Section III provides the method, the results, and a discussion of the results of the gesture elicitation study conducted while considering comfortable viewing positions; Section IV concludes the paper and suggests some future avenues to this work.

II. Related Work

A. Guessability and Elicitation

Eliciting actions from users to develop user-oriented systems stems from Good et al. [7], for creating a command-line interface. Guessability is a key factor in such symbolic input mechanisms, in which users use gestures or indicate characters or icons to perform actions on the system [8]. The *agreement rate* [4] measures the extent to which all participants agree on using a gesture for one command. Eliciting gestures from participants in order to obtain a user-defined referent-gesture mapping, was first carried out for surface computing [3]. Eliciting gestures from users would result in a usable system, where users remember their gestures better than those created by system designers [9]: three gesture sets (i.e., a user-defined set, a set designed by the authors, and a random gesture set) were compared by 33 participants and assessed with respect to *memorability* (i.e., how easy people can remember the gestures they have elicited). They evaluated the memorability immediately after creation, and one day after the elicitation. Participants remember the user-defined gestures much easier: up to 24% difference in recall rate compared to pre-designed gestures, people prefer user-defined gestures.

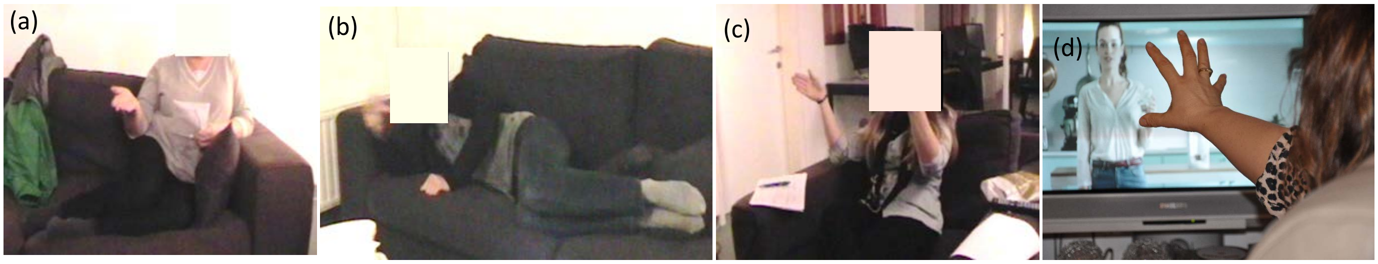

Fig. 1. Comfortable viewing positions elicited: (a) semi upright, (b) lying down, (c) sitting up right, (d) sitting right.

B. Gesture Elicitation Studies for Television

A gesture elicitation study was conducted for interacting by mid-air gestures with a smart television [10]. Another gesture elicitation study was conducted after with 12 computer science students for 12 common referents for TV interaction [11]. Furthermore, guidelines for the design of TV gesture sets were formulated. They found that one-hand gestures were heavily preferred to two hand gestures. In addition, participants use their fingers a lot requiring high resolution tracking equipment. They also make note of the observation that participants liked to draw letters in the air (e.g., C' for Close), and made use of WIMP-style interactions such as an air-push.

Since then, many more TV gesture elicitation studies have been conducted [5], [11]–[16]. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments showed that people prefer one-hand over two-hand gestures, and pointed to the importance of hand posture [11]. An iterative and more extensive approach to gesture elicitation was used [17]. Their user-centred approach included requirements gathering, functionality definition gesture elicitation, gesture design and usability evaluation. Free-hand and remote TV gestures were also compared [11]. Furthermore, they included a personalization feature for participants, and showed that the personalized gestures obtain better recognition rates than the standard gestures [17]. Another gesture elicitation study was conducted in the domain of the smart home [18]. Some of the referents derived during the study were also meant for TV interaction. After conducting a survey, defining the gesture vocabulary, studying the distinctness of the vocabulary, performing a memorability study and conducting a performance test, they indicate that users do not seem to differentiate between wider and smaller gestures, nor between different possible starting points of each gesture. Consequently, the recognition system should take this into account. As gesture recognition technologies get more accurate, and are increasingly more capable to recognize fine-grained actions [19], new measures are needed for these new possibilities. Fine-grained actions also incorporate the use of fingers [5], [12]: five measures were for Leap Motion gestures, and used these to evaluated the gestures elicited by their participants. A lower agreement rate was obtained compared to earlier studies because of the increased complexity and variability of unconstrained finger and hand gestures. In a study to investigate the performance an acceptance of gestures for TV with older adults, task completion time, error rate, usability and acceptance were measured [13]. The authors found that older people are generally positive towards gestures, and that the best implementation of gesture interaction is to directly transfer tracked hand movements to control a cursor on TV. They also investigated the effect of the device and the age [14]. Conflicts may also occur [16] as well as collaboration [15]. Finally, an extensive study on gesture elicitation for TV resulted in agreements scores exceeding those of existing studies [20]. The most recent study concerns a movable TV, but with another set of referents [21].

HeadGesture [22] consists of a Head Mounted Display (HMD) capturing simple gestures with head movement to interact with the devices so that users do not need to raise their arms to perform gestures or operate other remote controllers. Instead, they perform hands-free interaction. Similarly, head and shoulders gestures [23] offer the potential support of hands-free interaction, but these gestures are subject to physical fatigue more than other gestures. The question of replacing the classical TV remote control by gestures, whether they are mid-air or hands-free, remains an open question [17], [24], which requires further studies to determine the circumstances in which TV watchers would be akin to swap.

C. Qualitative and Quantitative Measures

In order to assess the results from gesture elicitation studies, we define the following set of qualitative and quantitative measures:

- Agreement rate: an indication for the extent to which participants agree on a gesture elicited [8], corrected [4].
- Memorability: a recall rate of a gesture some time after it has been elicited [9].
- Goodness-of-Fit: a subjective assessment rating the participant's confidence about how well the elicited gesture fits the referents [18].
- Ease of execution: a subjective assessment expressing the participant's perception of the ease of producing physically a gesture elicited [5].
- Ease of conception: a rating of how it is easy to come up with this gesture elicited for a referent [25].
- Enjoyability: same for the subjective perception of playfulness when eliciting a gesture [20].

The first gesture elicitation study for TV [12], further expanded in [5], is the most cited paper. Thus, it will be considered as a reference for comparing their results with respect to ours.

No.	Referent	Description	Average (15)
			BASIC REFERENTS (10)
1	TV on	Open the TV set	
2	TV off	Close the TV set	
3	Next channel	Go to next channel	
4	Previous channel	Go to previous channel	
5	Volume up	Increase the sound volume	
6	Volume down	Decrease the sound volume	
7	Volume mute	Turn off volume	
8	Open menu	Open the TV menu	
9	Hide menu	Close the TV menu	
10	Open item	Activate a menu item	
			GENERIC REFERENTS (2)
11	Yes	Provide an affirmative reply, confirm	
12	No	Provide a negative reply, cancel	
			TV-FEATURE REFERENTS (3)
13	TV Guide	Open the TV guide	
14	Play	Start playing a video or TV program	
15	Pause	Pause a video or TV program	

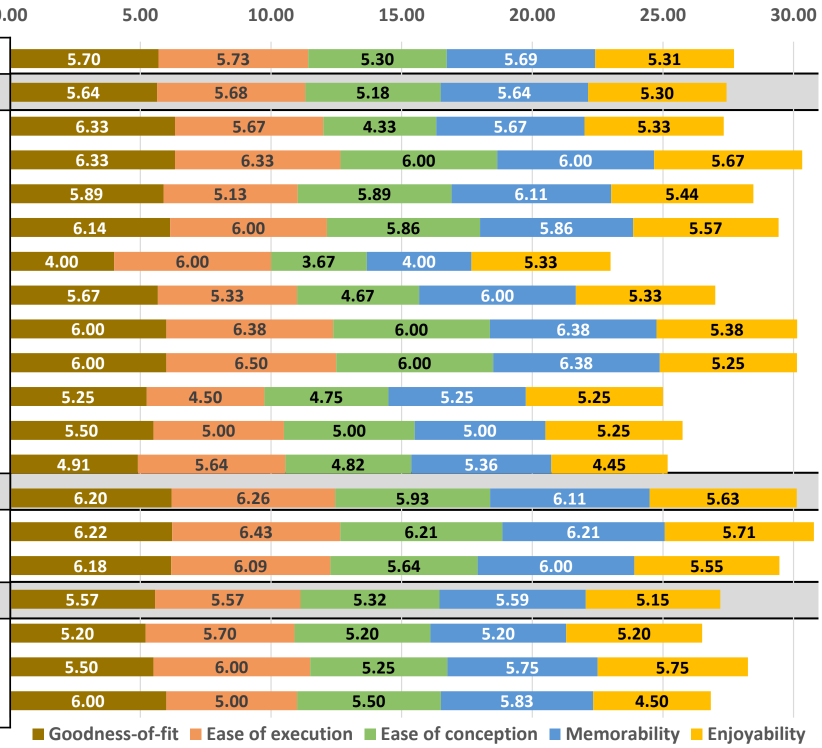

Fig. 2. Set of referents used in this study, their definition, and aggregated values for their measures.

D. Viewing Position

When we reviewed the state of the art, participants were asked to assume a viewing position that is rarely employed when watching TV. When people are watching TV at home, a leisure activity, they will more likely be in a comfortable position. An exception to this observation exists [26] in a elicitation study conducted for manipulating NetFlix and Spotify on a Smart TV while taking into account the physical position preferred by participants, but this study is specific to these two applications and their results are not necessarily generalizable. We argue that there is a need to investigate the influence of a natural, comfortable viewing position during gesture elicitation studies, which should produce a better ecological validity [6]. The recruitment in previous studies was often carried out at universities, involving students and/or staff. Consequently, the participants are relatively young (means between 20 and 30 years old), and often highly educated. Although recruiting more broadly from the general public might not result in less clear data, it increases the ecological validity [6].

III. User Study

A. Method

1) Participants: Participants were contacted via a recruitment agency with the following criteria: participants should be right-handed, a 50/50 male/female distribution, participants should watch TV regularly, and their age range should be broad. Participants brought a picture of their home room where they usually watch TV. The reason for this assignment was that by using this information we could form an accurate picture of the position of both the TV and participants when they are watching TV in their natural environment. This would help us to mimic the home situation and viewing position in the lab. A voucher worth 20 euros was provided as incentive and handed over after participation. Twenty-one participants (11 female, 10 male) were selected with an age between 18 and 65 years (M=41.4, SD=14.6) with diverse education degrees (i.e., secondary school, higher education, bachelor, university), a broad range of occupations (i.e., banker, manager, retiree, job seeker, student, lawyer, civil servant, nurse, housewife, self-employed, disabled), and in different living situations (i.e., single, married, divorced, living together, with/out children).

2) Stimuli: For our study we selected 15 referents common to TV interaction from current literature [5], [10], [13] divided into three groups (Fig. 2): a group of 10 *basic* TV commands (i.e., TV on/off", Next/previous channel", "Volume Up/down/mute", "Open/Hide menu", and "Open item"), a group of 2 *generic* commands (i.e., "Yes/no"), and a new group of 3 features (i.e., "TV Guide", "Play/pause").

3) Task: We welcomed the participants, briefed them about what they could expect, and, after signing an informed consent, invited them into the lab. Therefore, we discussed the pictures they brought along, which allowed us to mimic the home viewing position in our lab. Once participants assumed their comfortable, natural viewing position, we started with the study: in contrast to previous studies, we allowed participants to view all referents at once. They could revisit their gestures as any time. After participants indicated they completed gestures for all referents, recordings of their gestures were made. We believe that this could improve the results, which was suggested in [3]. Participants were given ample time to

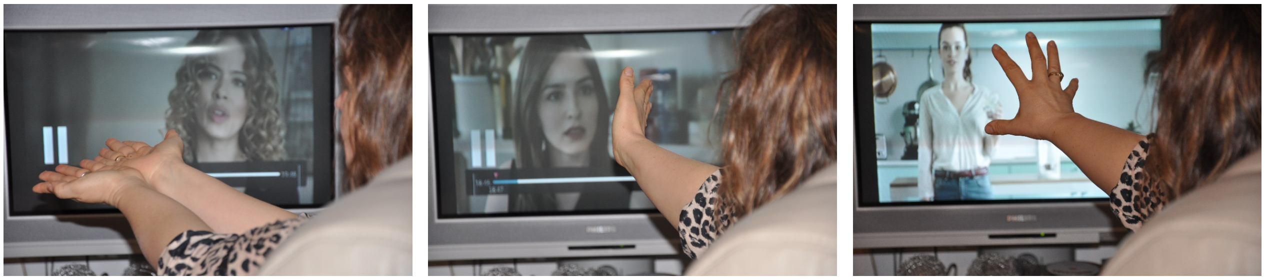

Fig. 3. Some sample gestures elicited for the Smart TV for a sitting person.

review the referents and were encouraged to try out several possibilities for each referent by the researcher who facilitated the study. Participants were given a pen (to write down descriptions, drawings, and anything that could help them in this exercise) and a list of the 15 referents (Fig. 2) created in three different versions with a different, randomized order of the referents. Opposite referents, such as `Volume up` and `Volume down`, were not necessarily placed consecutively. Each participant was free to elicit any set of samples wished (see some sample gestures in Fig. 3) and was also given a one-page questionnaire containing five statements to assess corresponding to the five aforementioned measures: *match*, *easy to perform*, *ease of conception*, *memorability*, and *fun*.

We alternated between a positive and a negative phrasing to ensure participants thoroughly read the statements and did not rush through the questionnaire. Participants rated these measures on a 7-point Likert scale, described their inspiration for the gesture, and whether performing the gesture resulted in any difficulty because of the viewing position.

B. Results

1) The Resulting Gesture Set: This set (Fig. 4) maximizes the agreement rate for all participants, calculated using the AGATe (AGreement Analysis Toolkit) software [4] for every referent, then aggregated by group, and for the whole set (M=.188, SD=.138). For the referents having almost equal agreement rates, the two most agreed gestures are provided (Table I). Overall, generic referents received a significantly better agreement (M=.42, SD=.087) than basic referents (M=.137, SD=.101) and TV-features (M=.139, SD=.091). Independently from these three categories, the elicited gestures roughly fall into three families depending on their agreement rate: high agreement ("Thumb up" for `Yes` with AR=.507, "Thumb down" for `No` with AR=.333, and "Point index" for `Open item` with AR=.323), medium agreement (ranging from "Palm up" for `Volume up` with AR=.293 and "Palm down" for `Volume down` with AR=.268 to "Swipe left/right" for `Next/Previous channel` both with AR=.216), and low agreement (ranging from "Open book" for `TV Guide` with AR=.091 to "Close book" for `TV Off` with AR=.043). Green bars in Fig. 4 denote the agreement rate obtained in this experiment with comfortable positions, while orange bars denote the corresponding agreement rate obtained in [12] with an upfront position. We can observe on Fig. 4 that rate obtained for these two different positions can be quite varying. There is no apparent correlation between agreement rates obtained for the two positions (Student's t-test returned t=.297, p=.77, *n.s.* – Levene's tests returned p=.850 by means, p=.603 by median, p=.851 by trimmed, *n.s.* – Q test returned q=.415, p=.771, *n.s.*).

2) Participants Viewing Positions: To examine each participant's position, three dimensions were distinguished (Table II): the position of the upper body, the feet, and how the back or the participant is oriented in relation to the couch. Seven participants were lying down; two of them on their right side, three of them on their left side, and two of them with their head or back against the left couch support. Six participants were in a semi upright position (Fig. 1a), meaning that they were not sitting fully upright, but rather in a more relaxed sitting or lying position where the upper body is more or less in a 45 degree angle (Fig. 1b). Three participants were leaning with their left arm on the left couch support, while their feet were in the couch oriented to the right. The three other participants in a semi upright position lay with their back against the back support of the couch and their feet on the coffee table. Six other participants were sitting upright in the couch (Fig. 1c). Two of them were leaning with their back against the left couch support, one was positioned with his back into the corner, and two participants were not leaning with their back (Fig. 1d). Participants were seated into either a larger sofa, or a seat for one person. This happened according to the furniture they normally occupy at home.

3) Results of the Post-test Questionnaire: Participants were asked to evaluate the goodness-of-fit of the gesture to the referent, the ease of execution, the ease of conception, the memorability, and the enjoyability to execute for each of their gestures after the test. Right part of Fig. 2 presents the results of these subjective measures for the gestures with the highest agreement rates. Participants evaluated these five aspects on a 7-point Likert scale, whereby 1 represents the worst possible score and 7 the best possible score.

To detect potential correlations between these measures, we computed Student's t-test to find out that there were only three correlations between all these responses:

- A highly significant correlation was found between Goodness-of-fit and Ease of conception (t=3.879, p^{**}=.002, Pearson's ρ=.720), thus suggesting that participants gave a good matching score to those gestures they found easy to create, design, retrieve, or invent. The easier an elicited gesture was to design, the better it is.

Referent	Description of the elicited gesture
`TV on`	(1) Point the index finger toward the TV, (2) Start from a fist, then open the hand with the palm toward the TV
`TV off`	(1) One hand is on the left side, the other at the right side, now bring them together horizontally, (2) Make a fist starting from an open hand pointed toward the TV
`Volume up`	Move the hand upwards with the palm of the hand facing up
`Volume down`	Move the hand from top to bottom with the palm of the hand facing down
`Mute`	(1) Move the hand from left to right, with the edge of the hand to the right, (2) Put the left hand on the left ear and the right hand on the right ear simultaneously
`Next channel`	Move the hand from right to left
`Previous channel`	Move the hand from left to right
`Open menu`	Start with both hands together, then move them apart horizontally, similar to opening the curtains
`Close menu`	One hand is on the left side, the other at the right side, bring both hands together horizontally
`Open Item`	Point the index finger toward the TV
`Yes`	(1) Hand toward the TV; (2) Make a fist with the thumb up
`No`	(1) Hand toward the TV; (2) Make a fist with the thumb down
`Play`	Make a clockwise circle with the hand
`Pause`	Make a cutting movement from top to bottom with the edge of the hand
`TV Guide`	Start with both hands against each other, turn both hands with the palm facing up, similar to a book that opens

TABLE I
GESTURES ELICITED FOR EACH REFERENT WITH THE HIGHEST AGREEMENT RATE. TWO GESTURES ARE GIVEN WHEN AGREEMENT RATES WERE EQUAL.

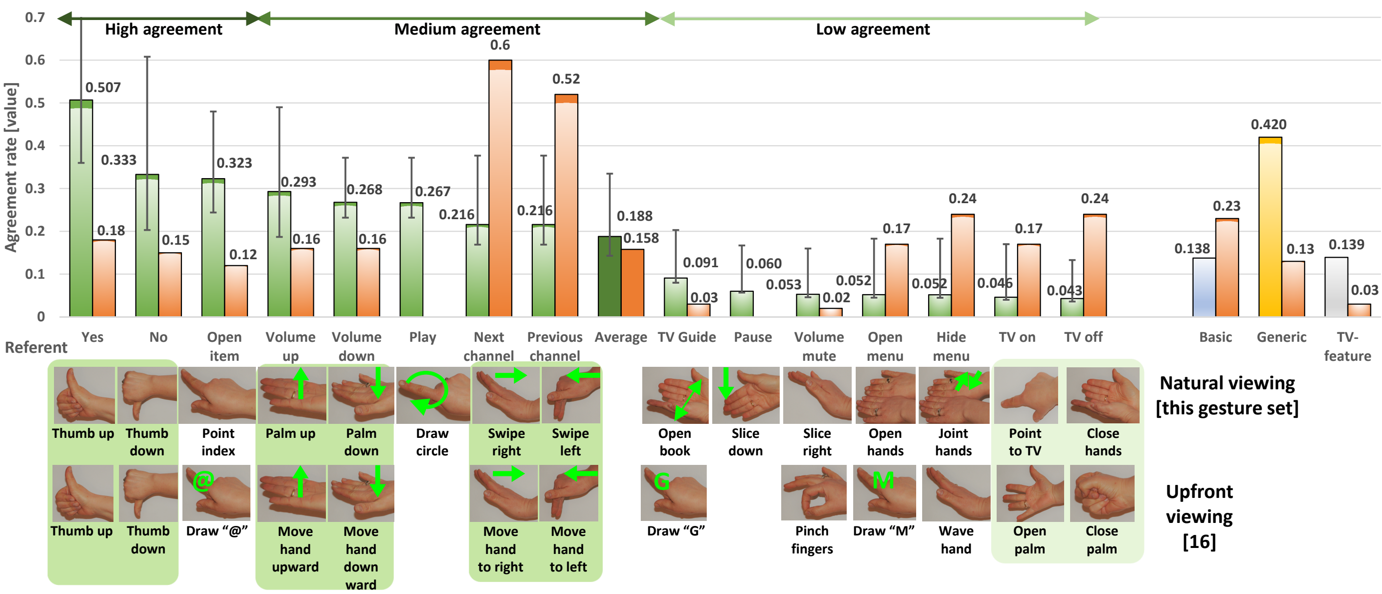

Fig. 4. Agreement rates for gestures proposed by participants: comfortable vs. upfront [4] with consensus gestures. Referents are ordered on the horizontal axis in descending order of their agreement rates for our study and by category; error bars show 95% confidence intervals. Green shadows highlight similar gestures across other studies [10]: darker green indicates a stronger confirmation than light green. When no such correspondence exists, no shadow is added.

- A very highly significant correlation was found between Goodness-of-fit and Memorability (t=6.212, p^{***}<.001, Pearson's ρ=.856), thus suggesting that participants also gave a good matching score to those gestures they remember easily. The easier they found a gesture to remember, the better the elicited gesture is.
- A very highly significant correlation was found between Ease of conception and Memorability (t=5.234, p^{***}<.001, Pearson's ρ=.813), thus suggesting that easily elicited gestures were also easier to remember.

4) Impact of the comfortable viewing position: In order to investigate what the impact of a natural, comfortable viewing position is on the creation and use of gesture for television interaction, we will first analyze the gestures for which participants indicated they had difficulties during execution. Note, that this alone will not suffice as, due to their position, participants might have elected for other gestures than they originally thought of using. In other words, their specific viewing position might have forced some participants to come up with another gesture. In the post-study questionnaires, participants were asked to indicate whether or not they experience any problems due to their specific position for each gesture they executed. The following bullets give an overview of the most significant issues reported by participants:

One participant complained about his ”hand being twisted” when performing the gesture for `Volume up`. In this case the gesture involved moving the right hand upward with the palm of the hand facing up. This participant was lying down on his left side. For ease of executing, he gave this gesture a score of three. Another participant lay down on his back with his head on the left support of the couch and wrote that performing his gestures for volume up and volume down can be ”a little bit difficult depending on the position”. After observing the footage, the back support of the couch might be

Upper body	Feet	Back
lying down	in the couch	back to the left couch support
lying down	in the couch	back to the left couch support
lying down	in the couch	left side
lying down	in the couch	left side
lying down	in the couch	left side
lying down	in the couch	right side
lying down	in the couch	right side
semi upright	in the couch	back to the left couch support
semi upright	in the couch	left side - in the corner
semi upright	in the couch	left side - in the corner
semi upright	on the coffee table	back to the back couch support
semi upright	on the coffee table	back to the back couch support
semi upright	on the coffee table	back to the left couch support
sitting upright	in the couch	back to the left couch support
sitting upright	in the couch	back to the left couch support
sitting upright	just outside the couch	back in the corner
sitting upright	on the ground	back loose from the couch support
sitting upright	on the ground	back loose from the couch support

TABLE II
OVERVIEW OF PARTICIPANTS' VIEWING POSITIONS, FIRST ORDER ON THE POSITION OF THE UPPER BODY, THEN THE FEET, AND FINALLY THE ORIENTATION OF THE BACK WITH RESPECT TO THE COUCH.

the cause: this person is making gestures with his right hand, which is also the side of the back couch support. Therefore, it might interfere with the performed gestures namely moving the arm up for volume up and moving the arm down for volume down. He made the same remark for `TV Guide`. This participant reported that performing his gesture for `Turn on TV` "required some effort". His gesture for `Turn off TV` involved making a fist (starting from an open hand) with his right hand. He wrote: "a bit unnatural when you are lying down, but given that you will be standing up [when you turn off the TV, perhaps to go to bed] it's ok." His assumption that you will stand up when turning off the TV is not always true.

One participant reported difficulties for the gesture to consult the TV guide. For this referent he opened his hands resembling a book that opens (this is also the gesture with the highest agreement rate for this referent, Fig. 4). Because he was lying down on his right side, his position blocked the movement of his right arm.

Again for the `TV guide` referent, difficulties were reported. A participant lying down on her back with her head on the left couch support, eventually opted to perform the gesture with her left hand. She reported: "no difficulties, but I had an earlier inclination to do this with the right hand". Important to note here is that all participants were right handed, and therefore this inclination might be quite natural for most participants. Similar to the previous participant, the fact that the back couch support might interfere with her right hand, caused her to execute the gesture with her left hand. She gave the gesture, a clockwise motion (so, the gesture itself is not too complicated) a 2 for ease of execution.

Two female, somewhat older participants, reported fatigue for many gestures. One participant said she had arthritis; the other required a long time to find and assume a comfortable position. If she did not find the right angle to lay backwards, she would start feeling pain in her neck, which reinforces the importance of conducting studies with a broad audience.

A younger participant was lying on his right side during elicitation. He reported difficulties to open his hand similar to a book that opens, due to being on right arm: "Because I lie down on my side I have one free arm. But I can execute it." He gave ease of execution a score of three and reported similar problems for `Open Menu`. For this referent, a participant, sitting right up with his back against the left couch support, reported that the back couch support interfered with his right arm. He scored ease of execution with 2.

A female participant lying down her back with her head on the left couch support, experienced problems when executing her gesture for `Previous channel`, which she gave a score of one for ease of execution. Her arm "felt strained" when moving her left hand to the left, with the thumb pointed to the left. A woman lying down on her left side felt "a heavy arm" when moving her hand upward with the thumb pointing up, her gesture for `Next channel` and the same for the opposite `Previous channel`.

A female participant chose to open her hands vertically for the referent `Turn on TV`. She was sitting upright with her back toward the left couch support. She experienced issues performing her gesture: "There is little room to bring my hand down". The score for ease of use was three.

A participant, lying down on his right side, changed from using his right arm to his left arm in order to perform the gesture for `TV off` (moving diagonally from the top right to the bottom left): "No, it is not really comfortable, perhaps it is better with the left arm."

Regarding the ease of execution assessed in the questionnaire (Fig. 2), the gestures having the highest agreement rates have a minimum score for ease of execution of 4.50 (M=5.69). This seems to contradict the results discussed in the previous paragraph, when only looking at the subjective scores for the gestures with the highest agreements scores would be enough for deciding a comfortable viewing position.

C. Discussion

1) Low agreement rates: Agreement rates for our study have a value lower than for other studies conducted with participants in an upright position [10] or mixed positions [27]. Is it due to the broader participants profile or because participants were positioned in a comfortable way, or both? There are some indications for the former. In the post-study questionnaire we also inquired about participants' inspiration for the gesture they conceived. The ideas on which participants based their gestures, were quite diverse: from characters in TV series, such as priests opening their arms during mass mess, to icons representing actions on digital equipment. A lot of other devices already offer similar functionality results in more diverse gestures. Participants referred to the Play and Pause icons on equipment, to swipe gestures and gesture to open menus on tablets, to mimicking button presses on the TV and on the remote control itself. This was certainly a factor in the great diversity of conceived gestures, which resulted in a lower agreement rate. The different viewing positions might also play some role. By looking at the many issues participants encountered during their elicitation and because of this sometimes changed their gestures, we believe that the various positions result in a more diverse gesture collection.

2) Impact of comfortable viewing position: A more diverse gesture set was obtained when taking into account the comfortable viewing position, which is expected to be more realistic. People who used their weaker hand found gesturing more difficult to execute. The highest agreement rate was found for gestures making use of two hands, which may cause problems for people lying down on one side, or seeking support with one arm on the side couch support. People lying on their back or sitting upright, with their back against the left couch support, have difficulties performing with their right hand and arm because the back arm support reduced the freedom of movement.

3) User Interface Conventions: We also saw that certain convention issues might have lowered the agreement rate. The most prominent example of this is the gesture set for `Next channel` and `Previous channel`. Eight participants chose "Swipe right" for the former referent, while four participants chose "Swipe right" for the latter referent. The opposite occurred with the "Swipe left" gesture. As their reported inspiration sometimes comes from touch screen interaction, this might be the cause of the lower agreement rate in this case. It is the case that for scrolling down on tablets you move down with your hand, whereas for scrolling down on computers you have to move the mouse of touchpad up (the latter can sometimes be altered in the settings). Another consequence of lying down is that those participants might not use the proper orientation for certain gestures. They are usually not lying down completely flat or sitting 'semi' upright in a perfect 45 degree angle; their positions, and the position of their heads more specifically, will be very diverse. The latter is important since we might use the position of our head to determine directions and angles for the gestures we perform. It is not certain that when people lying down on their left side with their head in a 20 degree angle, will be able to properly execute a gesture that requires horizontal movement. Because of these three issues (participants using their weaker hand, participants in viewing positions that partly obstruct their movement, and participants being in an awkward orientation), the error margin for the gesture recognition technology must be taken somewhat larger. Another suggestion for gesture recognition technology, which has been made elsewhere, is to offer options for personalization. Our 21 participants have taken on roughly 10 different positions. Furthermore, in households with many members, it is often the case that each member has his or her own typical spot in the living room, requiring a specific position in each case. Personalised gesture applications will certainly be of use in here.

4) Limitations: Although participants were positioned in a natural, comfortable viewing position, our lab still does not resemble the real TV viewing context. The position participants take at home was mirrored as much as possible, but there will still be the influence of the different furniture and the presence of other members in the household. Of course, the feasibility of conducting gesture elicitation studies at peoples' homes is not straightforward. Furthermore, our time was limited. Therefore, we were not able to include objective measurements on memorability. Methodologically, allowing participants to view all referents at once, and allowing them to revisit their gestures at any time is still recommended. However, we noticed that many participants found this a difficult exercise. Participants expressed both verbally and non-verbally that conceiving 15 gestures requires a substantial cognitive effort. Gesture recognition is still improving with more accuracy and can become more fine-grained [19]. We only compared our consensus gestures with those obtained in [10] (Fig. 4), but not with other studies [27]–[29] because they were produced under different experimental conditions, but this comparison could be achieved anyway.

5) Implementation: Once consensus gestures are identified among elicited ones, they need to be incorporated into the Smart TV device. Apart from using the API or SDK shipped with the Smart TV, other methods exist to support gesture interaction, such as by computer vision [29], by direct implementation into a wrapper [12], or by pattern matching [30].

IV. Conclusion and Future Work

We presented the results of our study into gesture elicitation for interacting with television sets. Our contribution to the state of the art lies in the way we gathered the data from participants. Whereas earlier studies placed participants in an upright position frontal to the TV, we allowed them to take a more natural position. More specifically, participants were taking place in the comfortable position they take when watching TV at home. As such, this study offers a greater ecological validity than the state of the art. With respect to the results we found lower agreement rates, likely due to the broader background of our participants and to the very diverse viewing positions used at home and in our elicitation study. The main problems of assuming realistic viewing positions is that people are sometimes forced to use their weaker hand, making the gesture execution less accurate and less easy; that people are obstructed in their movements because one arm or hand is blocked; and that some gestures that require two hands automatically cause problems for people how have on arm impaired due to their position. As this was only a first study into more realistic viewing positions for the use of gesture interaction for TV, we strongly believe that more field studies are required to gain a more complete understanding of realistic conditions and that accordingly adjustments to recognition technology should be explored and prototyped.

The next step of this gesture elicitation study consists in two actions: (1) feeding a new gesture recognizer with the consensus set resulting from this study. (2) extending the UsiXML UI description language [31] with the specification and mapping [32] of these gestures for supporting gesture user interface development and semi-automatic evaluation of guidelines based on them, similarly to [33], [34] for web usability guidelines.

References

[1] J. Vanattenhoven and D. Geerts, “Designing tv recommender interfaces for specific viewing experiences,” in *Proc. of TVX ’15*. New York, NY, USA: ACM, 2015, pp. 185–190. [Online]. Available: http://doi.acm.org/10.1145/2745197.2755522

[2] W. T. Freeman and C. D. Weissman, “Television control by hand gestures,” in *Proc. of IEEE Int. Workshop on Automatic Face and Gesture Recognition*. Piscataway USA: IEEE Computer Society Press, 1995, pp. 179–183.

[3] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, “User-defined gestures for surface computing,” in *Proc. of CHI ’09*. New York, NY, USA: ACM, 2009, pp. 1083–1092. [Online]. Available: http://doi.acm.org/10.1145/1518701.1518866

[4] R.-D. Vatavu and J. O. Wobbrock, “Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit,” in *Proc. of CHI ’15*. New York, NY, USA: ACM, 2015, pp. 1325–1334. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702223

[5] I.-A. Zaiţi, Ş.-G. Pentiuc, and R.-D. Vatavu, “On free-hand tv control: experimental results on user-elicited gestures with leap motion,” *Pers. and Ubiq. Comp.*, vol. 19, no. 5, pp. 821–838, Aug 2015.

[6] S. Kieffer, “ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations,” *AIS Transactions on Human-Computer Interaction*, vol. 9, no. 2, pp. 149–172, 2017. [Online]. Available: https://aisel.aisnet.org/thci/vol9/iss2/4

[7] M. D. Good, J. A. Whiteside, D. R. Wixon, and S. J. Jones, “Building a user-derived interface,” *Commun. ACM*, vol. 27, no. 10, pp. 1032–1043, Oct. 1984. [Online]. Available: http://doi.acm.org/10.1145/358274.358284

[8] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, “Maximizing the guessability of symbolic input,” in *Proc. of CHI EA ’05*. New York, NY, USA: ACM, 2005, pp. 1869–1872. [Online]. Available: http://doi.acm.org/10.1145/1056808.1057043

[9] M. A. Nacenta, Y. Kamber, Y. Qiang, and P. O. Kristensson, “Memorability of pre-designed and user-defined gesture sets,” in *Proc. of CHI ’13*. New York, NY, USA: ACM, 2013, pp. 1099–1108. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466142

[10] R.-D. Vatavu, “User-defined gestures for free-hand tv control,” in *Proc. of EuroITV ’12*. New York, NY, USA: ACM, 2012, pp. 45–48. [Online]. Available: http://doi.acm.org/10.1145/2325616.2325626

[11] R. Vatavu, “A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments,” *JAISE*, vol. 5, no. 2, pp. 187–211, 2013. [Online]. Available: https://doi.org/10.3233/AIS-130200

[12] R.-D. Vatavu and I.-A. Zaiti, “Leap gestures for tv: Insights from an elicitation study,” in *Proc. of TVX ’14*. New York, NY, USA: ACM, 2014, pp. 131–138. [Online]. Available: http://doi.acm.org/10.1145/2602299.2602316

[13] J. Bobeth, S. Schmehl, E. Kruijff, S. Deutsch, and M. Tscheligi, “Evaluating performance and acceptance of older adults using freehand gestures for tv menu control,” in *Proc. of EuroITV ’12*. New York, NY, USA: ACM, 2012, pp. 35–44. [Online]. Available: http://doi.acm.org/10.1145/2325616.2325625

[14] J. Bobeth, J. Schrammel, S. Deutsch, M. Klein, M. Drobics, C. Hochleitner, and M. Tscheligi, “Tablet, gestures, remote control?: Influence of age on performance and user experience with itv applications,” in *Proc. of TVX ’14*, 2014, pp. 139–146.

[15] O. Juhlin and E. Önnevall, “On the relation of ordinary gestures to tv screens: General lessons for the design of collaborative interactive techniques,” in *Proc. of CHI ’13*, 2013, pp. 919–930.

[16] K. Plaumann, D. Lehr, and E. Rukzio, “Who has the force?: Solving conflicts for multi user mid-air gestures for tvs,” in *Proc. of TVX ’16*. New York, NY, USA: ACM, 2016, pp. 25–29. [Online]. Available: http://doi.acm.org/10.1145/2932206.2932208

[17] H. Wu, J. Wang, and X. L. Zhang, “User-centered gesture development in tv viewing environment,” *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 733–760, Jan 2016. [Online]. Available: https://doi.org/10.1007/s11042-014-2323-5

[18] C. Kuhnel, T. Westermann, F. Hemmert, S. Kratz, A. Muller, and S. Moller, “I’m home: Defining and evaluating a gesture set for smart-home control,” *International Journal of Human-Computer Studies*, vol. 69, no. 11, pp. 693 – 704, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581911000668

[19] T. Sharp, C. Keskin, D. Robertson, J. Taylor, J. Shotton, D. Kim, C. Rhemann, I. Leichter, A. Vinnikov, Y. Wei, D. Freedman, P. Kohli, E. Krupka, A. Fitzgibbon, and S. Izadi, “Accurate, robust, and flexible real-time hand tracking,” in *Proc. of CHI ’15*. New York, NY, USA: ACM, 2015, pp. 3633–3642. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702179

[20] H. Dong, N. Figueroa, and A. El Saddik, “An elicitation study on gesture attitudes and preferences towards an interactive hand-gesture vocabulary,” in *Proc. of MM ’15*, 2015, pp. 999–1002.

[21] K. Stec and L. B. Larsen, “Gestures for controlling a moveable tv,” in *Proc. of TVX ’18*. New York, NY, USA: ACM, 2018, pp. 5–14. [Online]. Available: http://doi.acm.org/10.1145/3210825.3210831

[22] Y. Yan, C. Yu, X. Yi, and Y. Shi, “HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 198:1–198:23, Dec. 2018. [Online]. Available: http://doi.acm.org/10.1145/3287076

[23] J. Vanderdonckt, N. Magrofuoco, S. Kieffer, J. Pérez, Y. Rase, P. Roselli, and S. Villarreal, “Head and Shoulders Gestures: Exploring User-Defined Gestures with Upper Body,” in *Design, User Experience, and Usability. User Experience in Advanced Technological Environments*, ser. Lecture Notes in Computer Science, A. Marcus and W. Wang, Eds. Springer International Publishing, 2019, pp. 192–213.

[24] H. Wu, L. Yang, S. Fu, and X. L. Zhang, “Beyond remote control: Exploring natural gesture inputs for smart TV systems,” *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 4, pp. 335–354, Jan. 2019. [Online]. Available: https://content.iospress.com/articles/journal-of-ambient-intelligence-and-smart-environments/ais190528

[25] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum, “A procedure for developing intuitive and ergonomic gesture interfaces for hci,” in *Gesture-Based Communication in Human-Computer Interaction*, A. Camurri and G. Volpe, Eds. Berlin: Springer, 2004, pp. 409–420.

[26] R. Guérit, A. Cierro, J. Vanderdonckt, and J. L. Pérez-Medina, “Gesture Elicitation and Usability Testing for an Armband Interacting with Netflix and Spotify,” in *Information Technology and Systems*, ser. Advances in Intelligent Systems and Computing, . Rocha, C. Ferrs, and M. Paredes, Eds. Springer International Publishing, 2019, pp. 625–637.

[27] S. Ramis, F. J. Perales, C. Manresa-Yee, and A. Bibiloni, “Usability study of gestures to control a smart-tv,” in *Applications and Usability of Interactive TV*, M. J. Abásolo and R. Kulesza, Eds. Cham: Springer International Publishing, 2015, pp. 135–146.

[28] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and S. D. Marca, “Accelerometer-based gesture control for a design environment,” *Personal and Ubiquitous Computing*, vol. 10, no. 5, pp. 285–299, Aug 2006.

[29] M.-y. Chen, L. Mummert, P. Pillai, A. Hauptmann, and R. Sukthankar, “Controlling your tv with gestures,” in *Proceedings of the International Conference on Multimedia Information Retrieval*, ser. MIR ’10. New York, NY, USA: ACM, 2010, pp. 405–408. [Online]. Available: http://doi.acm.org/10.1145/1743384.1743453

[30] J. Vanderdonckt, P. Roselli, and J. L. Pérez-Medina, “!ftl, an articulation-invariant stroke gesture recognizer with controllable position, scale, and rotation invariances,” in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ser. ICMI ’18. New York, NY, USA: ACM, 2018, pp. 125–134. [Online]. Available: http://doi.acm.org/10.1145/3242969.3243032

[31] Q. Limbourg, J. Vanderdonckt, B. Michotte, L. Bouillon, and M. Florins, “Usixml: A user interface description language supporting multiple levels of independence,” in *Proc. of ICWE’2004 Workshops*. Berlin: Springer, 2004, pp. 325–338.

[32] F. Montero, V. López-Jaquero, J. Vanderdonckt, P. González, M. Lozano, and Q. Limbourg, “Solving the mapping problem in user interface design by seamless integration in idealxml,” in *Interactive Systems. Design, Specification, and Verification*, S. W. Gilroy and M. D. Harrison, Eds. Berlin, Heidelberg: Springer, 2006, pp. 161–172.

[33] J. Vanderdonckt, A. Beirekdar, and M. Noirhomme-Fraiture, “Automated evaluation of web usability and accessibility by guideline review,” in *Web Engineering - 4th International Conference, ICWE 2004, Munich, Germany, July 26-30, 2004, Proceedings*, ser. Lecture Notes in Computer Science, N. Koch, P. Fraternali, and M. Wirsing, Eds., vol. 3140. Springer, 2004, pp. 17–30. [Online]. Available: https://doi.org/10.1007/978-3-540-27834-4_4

[34] J. Vanderdonckt and A. Beirekdar, “Automated web evaluation by guideline review,” *J. Web Eng.*, vol. 4, no. 2, pp. 102–117, 2005. [Online]. Available: http://www.rintonpress.com/xjwe4/jwe-4-2/102-117.pdf