User-Defined Gestures for Free-Hand TV Control

Radu-Daniel Vatavu
University Stefan cel Mare of Suceava
13 Universitatii
720229 Suceava, Romania
vatavu@eed.usv.ro

ABSTRACT
As researchers and industry alike are proposing TV interfaces that use gestures in their designs, understanding users' preferences for gesture commands becomes an important problem. However, no rules or guidelines currently exist to assist designers and practitioners of such interfaces. The paper presents the results of the first study investigating users' preferences for free-hand gestures when controlling the TV set. By conducting an agreement analysis on user-elicited gestures, a set of gesture commands is proposed for basic TV control tasks. Also, guidelines and recommendations issued from observed user behavior are provided to assist practitioners interested in prototyping free-hand gestural designs for the interactive TV.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: User Interfaces. Evaluation/methodology, user-centered design.

General Terms
Design, Experimentation, Human Factors.

Keywords
Gestures, gesture recognition, free-hand, Kinect, TV, living room, experiment, study, guessability, user-defined, interactive TV.

1. INTRODUCTION
Following the success of motion-sensitive controllers for video games (Wii/Kinect), gestural interfaces are slowly but steadily invading the rest of the controllable living room. Researchers and industry alike have been experimenting interface designs that allow users to control the functions of their TV sets by means of gestures only [2,4,7,12,13]. Such designs either augment the TV remote control with motion sensing capabilities [2,15]; replace the remote with smart mobile phones [7]; reuse video game controllers [13]; appropriate pieces of furniture as supporting equipments [12]; or employ free-hand gestures [4,5].

There are many good reasons for adopting gestures for TV control. Among the most often mentioned ones [2,12], gestures come to replace mis-designed remote controls that affect negatively the user experience. For example, ethnographic studies [3] have showed numerous times that many household members find remote controls complex and difficult to use with the main reasons being small buttons and high button congestions. Also, despite the large available palette of functionalities, people only use "*selected buttons of [their] remote controls … [because] some are really useless*" [3]. Other weak points are the lack of backlighting on most remotes which reduces their usability in the dark; the fact that remotes can be lost; and their periodical need for fixing and battery replacement. On the other side, gestures feel familiar and natural to their users when properly designed while they also bring a new dimension of *fun* into the interaction (heavily exploited by today's gesture-sensitive video game controllers). However, probably the most important advantage of using gestures lies with the possibility of customization: users can provide their own commands without being constrained by the rigid design of the remote control.

Despite such advantages, we must also note that today's gestural interfaces are far from being ideal [9]. Although marketed as natural user interfaces, gestures have side effects such as misrecognition and poor designs of the gesture-to-function mapping [9]. Free-hand motions also induce fatigue with prolonged usage. However, as practitioners will learn more about their users in terms of execution capabilities and mental models for mapping gestures to tasks, gestural interfaces are likely to conquer the future living-room environment (either on their own [4] or as part of mixed designs [2,13,15]).

This paper contributes on the little researched area of designing intuitive mappings between gestures and TV functions. We report results from a study addressing users' preferences for free-hand gestures when controlling the TV set. We compile a set of gestures for basic TV commands and provide guidelines for practitioners interested in TV gestural designs. In this context, this work represents the first exploration of users' preferences for free-hand gestures for the interactive TV.

2. RELATED WORK
Researchers have always shown a strong interest in designing new interactions for the living room. An early prototype of Freeman and Weissman [4] employed a video camera which detected the location and posture of the viewer's hand for performing click-like interactions on the TV. The ARGUS system [5] extended the concept to multiple cameras employed to detect the direction of pointing in order to allows users to control multiple equipments with the same gestures. Vatavu and Pentiuc [12] showed how TV gestural interfaces can be developed robustly and without privacy concerns with the interactive coffee table prototype. WIMP-style interactions (Windows, Icons, Menu, and Pointing) were also explored for multi-screen home entertainment systems [13]. Such systems exploit the experience people had already acquired from working with windows-based operating systems in order to provide a smooth transition to other interactive spaces.

In the measure that gesture acquisition devices are becoming affordable (such as low-cost Wii/Kinect sensors), the main design problems are shifting towards finding the optimum mapping from gesture to function [11,14]. However, existing toolkits for gesture design only advise practitioners on how to obtain a high

EuroITV'12, July 4-6, 2012, Berlin, Germany.

recognition rate or help them organizing the gesture set [1]. As a consequence, the only options for finding *good* gesture mappings remain to rely on the designer's own expertise or to organize user studies. Norman [9] showed how the first option has often led to mis-designs and frustrating user experiences.

Involving users into the design process represents a viable alternative for collecting important data to inform design. In this direction, Lessiter et al. [8] employed users to suggest labels for the buttons of the remote and to rank their intuitiveness. Wobbrock et al. [14] introduced a methodology for eliciting gesture commands from users in order to compile sets of high-agreement gestures. Follow-up works verified the methodology for mobile phones [7,11], public displays, and tabletops [6]. However, no study yet exists for investigating user preferences for free-hand gestures despite many gestural interfaces being proposed in the TV community. This work tries to fill this gap by collecting and reporting user-elicited free-hand gestures to inform practitioners of the next generation TV gestural interfaces.

3. EXPERIMENT

An experiment was conducted in order to elicit user preferences for free-hand gestures for frequent TV control tasks. The experiment was designed and ran in accordance with domain practices for such guessability studies [11,14].

Participants

12 participants volunteered for the study (mean age 21 years, all right-handed, 6 females). Although volunteers were CS students, none of them had previous experience in interaction design nor had they used gestural interfaces before (such as Kinect).

Apparatus

Gestures were acquired using the Microsoft Kinect sensor[1] which was connected to a 3.0GHz computer running a custom software application which directed the experiment. The application logged video recordings and body tracking data of participants performing gestures. The Kinect sensor allows for 20 body points to be tracked in real time at about 30 frames per second. The PC was connected to a video-projected display (with a size of 2×1.5 meters and a resolution of 1024×768 pixels) which was emulating a large TV equipment. A sufficient amount of free space was secured for the testing area in order for participants to perform gestures in a confident and safe manner.

Referents

12 commands commonly used for controlling the TV set were selected for the experiment (see Table 1). By following the terminology of [14], commands are denoted as *referents*.

The set of referents contains frequently-used commands for controlling the TV set which cover channel navigation, volume control, accessing extra options through menus, performing general enquiries, and answering system questions. Although other commands could have been easily included, we opted for this specific and limited set for several reasons. First, ethnographic studies showed that participants focus and use few buttons only [3] and, when asked, redesign their own versions of the remotes with solely the functions they actually use or need. Second, the number of gesture commands people can remember for effective use should be limited in order not to increase cognitive load. More individual commands would translate into complex designs with a similar puzzling effect exhibited by today's remote controls [3]. This very observation has led the industry in proposing short versions of the remote such as the Tek Pal Simple Remote[2] or the Flipper[3] featuring 6 frequently-used buttons only (power, channel +/-, volume +/-, mute), all of which are part of our set. Also, menus represent a viable option to group other, less frequent, commands [2]. We need to note that the set can be easily enlarged to include air-drawn digits 0 to 9 (for quick channel change commands or similar short-cuts). For obvious reasons there was no point including digits in this study.

For each referent, a video demonstration depicting its effect was prepared in advance to be displayed during the experiment.

Table 1. Referents used for the elicitation experiment

REFERENT		NOTES
1.	SELECT/OPEN	Open TV or indicate a specific TV screen in a multi-screen environment
2.	CLOSE	Close the TV screen
3.	NEXT	Go to next channel
4.	PREVIOUS	Go to previous channel
5.	VOLUME +	Volume control
6.	VOLUME -	
7.	MUTE	
9.	OPEN MENU	Pop up contextual menu
9.	HIDE MENU	Hide currently opened menu
10.	HELP	Ask system for help
11.	YES	Answer to system question / Accept
12.	NO	Answer to system question / Reject

Procedure

Participants were asked to propose free-hand gestures using one or both hands for each referent in the set. Each referent was presented using a short text description (e.g. *volume up*) and by running a video demonstration of its effect. After participants confirmed they understood the effect, they were asked to suggest a suitable gesture command that would trigger the effect they had witnessed. Participants were encouraged to think aloud while searching for the best gesture to match the task. No visual feedback was provided during the gesture elicitation phase. The order of referents was randomized across participants.

4. RESULTS

We report results from 12 (participants) x 12 (referents) = 144 proposals for gesture commands. An agreement analysis was conducted in order to understand participants' level of consensus for each specific task/referent. The agreement rate measure of Wobbrock et al. [14] was used for this purpose:

$$A_r = \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)^2$$

where P_r represents the set of all gestures proposed for referent r and P_i represent groups of similar commands. The agreement rate A_r ranges from $|P_r|^{-1}$ (no agreement at all) to 1 (perfect agreement).

The mean agreement for our set was $.42$ ($SD = .28$), comparable to values obtained by [14] ($.32/.28$) and [11] using the same methodology for surfaces and mobile phones. Individual rates for each task are shown in Figure 1.

[1] Kinect for Windows, http://kinectforwindows.com

[2] TekPal Remote, http://www.bigbuttonremotes.com/remotes-tekpal.htm

[3] Flipper Remote, http://www.flipperremote.com/

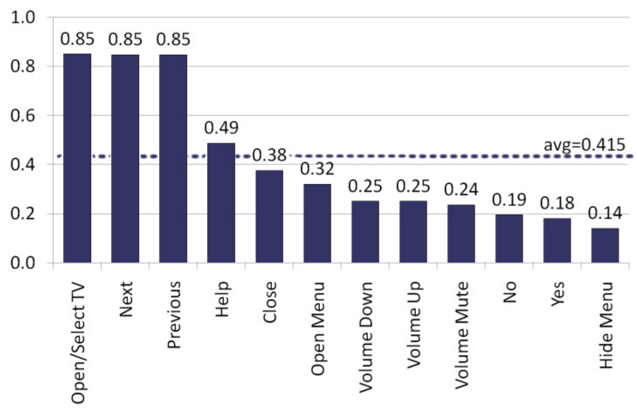


Figure 1. Agreement rates for user-defined free-hand gestures

Participants' gestures consisted in both motion and hand posture data. For some tasks (e.g. *Next*), the motion component was the most relevant whilst hand posture was just an artifact of the motion (i.e. non-intended). Other tasks clearly relied on hand posture (e.g. *Volume mute)* whilst hand location and motion was irrelevant. We consistently applied these rules when calculating the agreement rates. The highest agreement ($.85$) was obtained for *Open*, *Next*, and *Previous* (consensus among 11 out of 12 participants). The smallest levels of agreement ($< .20$) occurred for *No*, *Yes*, and *Hide menu*. In order to better understand what these values actually represent and extract meaningful information, we illustrate a few examples of how they could be obtained. For example, an $.50$ agreement level can be obtained when two groups of six participants propose the same command or when two commands are supported by two groups of 8 and 4 (suggesting a robust agreement):

$$A_r = \left(\frac{6}{12}\right)^2 + \left(\frac{6}{12}\right)^2 = .50; A_r = \left(\frac{8}{12}\right)^2 + \left(\frac{4}{12}\right)^2 = .55$$

An agreement rate of approximately $.25$ can be attained with four groups of participants each supporting a different command or with a group of at least 6 in consensus:

$$A_r = 4 \times \left(\frac{3}{12}\right)^2 = .25; A_r = \left(\frac{6}{12}\right)^2 + 6 \times \left(\frac{1}{12}\right)^2 = .29$$

Therefore, agreement rates above $.50$ are indicators of robust proposals; agreements above $.25$ suggest useful commands; while values lower than $.20$ indicate the need of expert design. In total, 4 referents had agreements over $.50$ (33%), $9\ (= 4 + 5)$ over $.25$ (66%), and 3 were under $.20$ (25%).

4.1 A gesture set for free-hand TV control

The agreement analysis justifies an attempt to compile a gesture set for TV control. When performing the final gesture-function associations, the gesture occurring most frequently for a referent won the command. Figure 2 illustrates the final proposal.

Open/Select TV was performed by pointing towards the screen and confirming the command with either a hand posture or a quick push of a virtual air button (a technique we denote as *air-push*) by 11 out of 12 participants. In order to perform *Close*, participants drew letter *X* in mid air (mimicking the close icon of windows on PCs). Two distinct gestures were included in the set for *Open Menu*: letter *M* (4/12) and a double *air-push* (5/12). Low agreements were found for *Hide Menu* (0.14) for which participants suggested different options (*X* for closing; another *M*; *C* for close; move menu away with the hand; shake hand rapidly to ignore, etc.) Therefore, an artificial choice was made for *Hide*: the same gesture as for *Open Menu* inspired by the design of existing remote controls that only have a single menu button. *Next* and *Previous* were performed by moving the dominant hand to left and right (11/12).

Controlling the volume revealed interesting findings. *Volume Up* and *Down* exploited the use of both hands as well as the distance between hands as a control parameter. While the non-dominant (left) hand was held fixed, the right hand moved along one axis to control volume level (8/12). A considerable high (and unexpected) agreement was found for *Mute* despite its abstract nature. Almost half (5/12) participants performed a hand gesture by closing fingers simulating a closing mouth (shut up). In order to invoke system help, the majority of participants (8/12) drew a question mark in mid-air. *Yes* and *No* received culturally-dependent gestures such as thumbs-up (4/12), thumbs-down (4/12), and vertical/horizontal movements of the hand (2/12).

4.2 Guidelines for TV gesture set designs

Observations collected during the experiment and the follow-up analysis were compiled into a set of guidelines for practitioners:

(1) We found that one-hand gestures were heavily preferred to two hands (10 out of 12 commands use one hand). Two hands served for parameterized tasks (e.g. volume level). The same reasoning could therefore be reused for other similar situations (such as brightness or contrast adjustment);

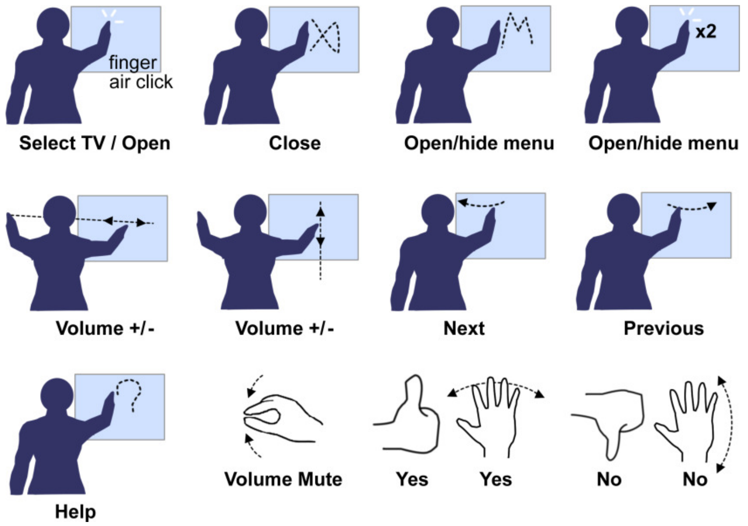
(2) Hand posture was found to be important. The index finger pointed occurred most frequently but other postures such as thumbs-up, thumbs-down, and hand closed were equally employed. One participant performed *Close* by closing the fingers of her hand. Another participant used 3 stretched fingers in order to pop up a contextual menu (*Open Menu*). This shows the need of sensing technology with high enough resolution to capture such details (an important note for ambient video cameras);

(3) A clear preference was observed for drawing letters in mid-air (*X*, *M*, and question mark are present in the set; *Y* and *O* were suggested for yes/ok; *N* for no; *C* for close, *H* for help, etc.). This suggests possible extensions to other letter-like commands with the 0..9 digits falling into this category. Participants witnessed letters were easy to use and suspected they would also be easy to remember;

(4) Pointing was frequently used and WIMP-style interactions could be observed in many cases. For example, *air-push* simulates clicking a button while volume control gestures remind the use of track bars. An analogy with a long mouse click was used by a participant who executed *Open* with a long pointing (~2 sec) towards the screen. One participant imagined popping up a volume control (similar to the ones found in PC media players) by drawing letter *V*. Some preferences were observed for tangible controls. For example, one participant imagined a turning button in mid-air to be used for controlling the volume (wrist movements to right/left would change the volume level up/down);

(5) Cultural gestures were proposed due to their similarity with existing situations in the real-world (*Yes/No*). Although their generality is limited from culture to culture, their use should be exploited due to familiarity to their users [10];

(6) Some TV functions may be too abstract for participants to suggest meaningful gestures or to remember special designed ones. For such cases, on-screen widgets and menus should be used. This observation could also raise questions on the appropriateness of a gestural interface for some TV control tasks for which a mixed design (remote + gestures) might prove more appropriate;



REFERENT	GESTURE
Select/Open	Point + air-push/air-click
Close	Draw letter X
Next	Move dominant hand to the left
Previous	Move dominant hand to the right
Volume +	Non-dominant (left) hand serves as the reference the while right hand moves left/right or up/down
Volume -	
Mute	Hand closes the fingers (shut-up)
Open Menu	Draw letter M or double air-push
Hide Menu	Draw letter M or double air-push [gesture artificially created by copying Open Menu]
Help	Draw question mark
Yes	Thumbs-up / shake horizontally
No	Thumbs-down / shake vertically

Figure 2. Gesture set proposed for frequent TV commands as compiled from the participants' agreement rate analysis.

(7) Although this study only investigated free-hand gestures, body-referenced gestures could be further explored. They occurred for *Volume Mute* for which one participant moved both hands to the ears covering them while another brought the hand close to her mouth. Also, one participant lifted his shoulders in order to invoke *Help*.

Cause of limited space we provide the complete set of gesture suggestions as a download from http://www.eed.usv.ro/~vatavu

5. CONCLUSION

The paper presented the results of the first study addressing users' free-hand gesture preferences for TV control. A set of candidate gestures was proposed for 12 common TV control tasks. We hope the results will benefit practitioners interested in designing gesture-based interfaces for the next generation of home entertainment systems and smart living room environments. We equally hope that the observations of this study will also raise questions on the usability of gestural interfaces for TV control, at least for some abstract tasks. Future work will investigate users' preferences for remotes vs. free-hand gestures.

6. ACKNOWLEDGMENTS

This paper was supported by the project "Progress and development through post-doctoral research and innovation in engineering and applied sciences- PRiDE - Contract no. POSDRU/89/1.5/S/57083", project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

7. REFERENCES

[1] Ashbrook, D., Starner, T. 2010. MAGIC: a motion gesture design tool. In Proc. of CHI '10. ACM, New York, NY, USA, 2159-2168

[2] Bailly, G., Vo, D.B., Lecolinet, E., Guiard, Y. 2011. Gesture-aware remote controls: guidelines and interaction technique. In Proc. of ICMI '11. ACM, New York, NY, USA, 263-270

[3] 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. Computers in Entertainment. 6 (1), Art 5 (May 2008)

[4] Freeman, W.T., Weissman, C.D. 1995. Television Control by Hand Gestures. In: Proc. of IEEE Int. Workshop on Automatic Face and Gesture Recognition, Zurich

[5] Kohler, M. 1998. Special topics of gesture recognition applied in intelligent home environments. In: GW'1997. LNCS, vol. 1471, 285-296. Springer, Heidelberg

[6] Kray, C., Nesbitt, D., Dawson, J., Rohs, M. 2010. User-defined gestures for connecting mobile phones, public displays, and tabletops. In Proc. of MobileHCI '10. ACM, New York, NY, USA, 239-248

[7] Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S. 2011. I'm home: Defining and evaluating a gesture set for smart-home control. International Journal of Human-Computer Studies, 69 (11), 693-704

[8] Lessiter, J., Freeman, J., Davis, R., Dumbreck, A. 2003. Helping Viewers Press the Right Buttons: Generating Intuitive Labels for Digital Terrestrial TV remote controls. Psychology, 1(3), 355-377

[9] Norman, D.A. 2010 Natural user interfaces are not natural. Interactions 17, 3 (May 2010), 6-10

[10] Rehm, M., Bee, N., André, E. 2008. Wave like an Egyptian: accelerometer based gesture recognition for culture specific interactions. BCS-HCI'08, British Computer Society, 13-22

[11] Ruiz, J., Li, Y., Lank, E. 2011. User-defined motion gestures for mobile interaction. CHI '11. ACM, 197-206

[12] Vatavu, R.D., Pentiuc, S.G. 2008. Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience. In Proc. of EUROITV'08, Springer-Verlag, Berlin, Heidelberg, 183-187

[13] Vatavu, R.D. 2011. Point & Click Mediated Interactions for Large Home Entertainment Displays, Multimedia Tools and Applications, Springer, doi:10.1007/s11042-010-0698-5

[14] Wobbrock, J.O., Morris, M.R., Wilson, A.D. 2009. User-defined gestures for surface computing. In Proc. of CHI '09. ACM, New York, NY, USA, 1083-1092

[15] Smart TV by LG, http://www.lg.com/us/tv-audio-video/discoverlgtvs/smarttv/index.jsp