

Specification and Usability Evaluation of an Attention-aware Remote Control designed in a Physical Prototype

Daniel Chagas

LUQS

Universidade de Fortaleza

Fortaleza, Brazil

prof.daniel.chagas@gmail.com

Elizabeth Furtado

LUQS

Universidade de Fortaleza

Fortaleza, Brazil

elizabethsfur@gmail.com

Abstract— This paper defines requirements for entering text in interactive digital TV, based on theories of shared attention, in order to make a prototype of a remote control which enables more natural user interaction. The physical prototype of this newly created device features movement recognition and sensory feedback as modalities of interaction. In usability tests, data on users' performance and satisfaction was collected, as well as data on their cognitive load (attention) and state of meditation (relaxation) captured through an Electroencephalogram device. The results showed that the solution, analyzed for a sample of 18 users, increased performance for typing long texts by 26.5%, raised satisfaction scores by 15%, relaxation scores by 29.4%, and maintained the users' cognitive load when compared to use of an infrared remote control.

Keywords— *Alternative Remote Control; interactive TV; Attention-aware System; Arduino.*

I. INTRODUCTION

The evolution of digital television and its connectivity capabilities are creating new possibilities for viewers to interact with television content. Bachmeyer [1] affirms that there are two trends in interactive digital TeleVision (iTV): the social iTV, where the viewer uses his/her iTV and its programs to socially interact through text; and the collaborative iTV, where the viewer uses his/her iTV to create new video content. However, both iTV trends present the common problem of data input. In [2] researchers show that usability and accessibility issues in iTV application are derived from standard Remote Control (RCs). Researches with standard infrared RC show its inefficiency when texting large amounts ([2], [3]) and open precedent for developing new ways so the user interact with this device.

In the UK, in a survey by Cooper [4] reporting user experiences with iTV system, shows that the most common solutions for typing texts are the virtual keyboard, multi-tap keyboard, and the word prediction systems. Solutions engaging for full keyboards embedded in RCs are mentioned in the study [5]. Several other studies, using voice [3], [5], gestures [6], touch in screen devices [7], or a combination of these modalities of interaction [8] proved to be either inefficient for large text input, or too costly for the mainstream market. These solutions present problems linked to human factors (fatigue and cognitive effort) during long and simultaneous tasks. Human cognitive

factors must be considered during the development of physical devices in order to assure comfortable user experiences with iTV applications [9].

While we advocate for a combination of different ways of interaction [10], we also considered a well-established (but insufficient) paradigm of interaction: the RC, that was treated here as an alternative for entering long texts in TV by making some improvements in its physical design specification and interactive language. Our decisions for such improvements were conducted by studies in human factors [11]. So, in the RC context of use, an important example of human factor that affects the inefficiency in large text input is regarded to the human attention. Attention is defined as the set of mechanisms which allow for the allocation of cognitive resources, often times limited [12] (attention as selection paradigm).

Existing RC solutions require that the user's attention switches between the input device, the TV (the iTV applications) and the content (television programs). Figure 1 illustrates that in the context of the infrared RC manipulation for TV data input, using the virtual keyboard (most common modality in TV), the user is in a situation of shared attention between the tasks of thinking about on what to type (1), in manipulating the RC (for look at the keys (2 and 4) and pressing it (3)), and checking the result on the TV screen (5). An alternating flow of attention between the user and the TV impairs any attempt at natural interaction. Systems that take into account the attention allocation are called attention-aware systems [10].

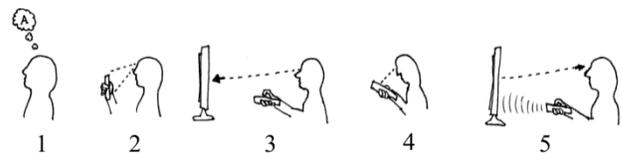


Fig. 1 - Switching one's attention while using an Infrared RC

This research was aimed at specifying, prototyping and evaluating a RC solution, which we named MoveRC, in order to make user interaction more natural when entering text in TV. The RC we've proposed uses a combination of ways of interaction, including language commands by movements in

order to maximize the automaticity of the input process data in TV as soon as a screen to minimize effort during the interaction.

After the description of hypotheses and related work, we describe the theoretical basis, the requirements relating to human factors, the product design, as well as the results of using this type of RC by users compared to using an infrared RC. The main points and challenges are presented before the conclusion section.

II. HYPOTHESES OF THIS RESEARCH

The challenges that we had in order to identify the RC as an alternative for entering long texts are associated with the characteristics of infrared RCs which affect the physical and cognitive effort of users. Such characteristics were identified based on experiences of observation of users using RCs to perform the task of inputting text on the TV [11] and consist of:

i) Lack of feedback: Feedback of the actions requiring answers from users are provided on the TV screen, forcing the viewer to divide his/her attention between controlling the RC and checking the response on the TV. The first challenge was to define how tasks of user interaction with the TV (such as recognition, selection, feedback, etc.) should be allocated on different TV artifacts (such as iTV applications, the TV set itself, and the RC); and

ii) Inappropriate ergonomics for interactivity: Slightly tilting the control or lifting one's arm so the user is able to maintain a straight line from the infrared RC to the TV in order to interact are tasks that often block users from seeing which button they're pressing. The second challenge was to minimize user fatigue in the effort to push the buttons while maintaining a line-of-sight between the RC and the TV.

We hypothesized that the use of a device for data input that meets requirements related to user attention and ergonomics can improve user interaction when inputting text on the TV. The secondary hypotheses are related to the use of a device for data input while meeting those requirements: i) Does it improve user performance in inputting text on the TV? ii) How can the continuous use of such device improve user performance? iii) Does it improve user satisfaction in interacting with the TV? iv) Does it improve the allocation of attention in the user's interaction with the TV?

III. RELATED WORK

A. Modalities of Interaction

In the following article several interaction modalities were identified, which were implemented in solutions for iTV systems in order to develop interactive activities of recognition, selection, feedback, etc. The modality aspect is associated to human senses (such as vision, speaking, hearing, and touch). The solutions implemented are:

- **Point-and-click:** this solution is typical for a pointing RC, which uses movements. It is commercially implemented by companies such as LG, Samsung and Nintendo, and consists of an RC, with resources that enable spatial mapping of the device and allow the user to move a specific "target" on the screen;

- **Voice Recognition for television:** refers to the usage of commands which are pre-set to operate the television;
- **Motion recognition:** usually refers to the usage of a built-in camera in the TV device which enables the recognition of the user's hand palm which controls a cursor on screen;
- **Body parts used as device:** refers to using any body part (such as hand palm) as a mean to input television commands; and
- **Usage of a second screen.** The usage of the TV device in combination with another device, which features touch screen technology, such as tablets or smartphones.

B. Academic and Commercial Solutions

The purpose of this section was to present innovative solutions in iTV systems for user interaction, originated from researches developed by Academia and Private Market.

The pointing RCs make use of several technologies (gyroscopes and accelerometers) to move a cursor in a more natural manner, by transposing the user's hand movements to the screen. Through that feature, not only are TV basic functions accessible from RC buttons, but the user's hand movements also allow for the entry of texts due to the point and select concept. To point at letters and buttons on screen allows the user to have adequate agility and speed when entering short texts. However, the ergonomics of having to maintain the control constantly pointed at the television discourages use for longer periods when it comes to texting.

Voice recognition presents great performance in pre-defined activities, which use pre-set key words, and in quiet environments [13]. However, its use in free text entry demands, not only high computational power, but also user's mindfulness toward pronunciation, spelling and tone and a system capable of interpreting accents and regionalism. However, in analyzes by specialized magazines these features proved to be impractical, as they are slow and exhausting.

In [14] the authors used a Wiimote, the Wii videogame control, to compare the efficiency of pointing tasks at the screen. The LG Company has also created a simplified control that uses movements. However, neither of the solutions is focused on data input, but on the point-and-click interaction. Scholar researches on gestures and iTV are quite common in academia. Many of them try to replace the RC, as PalmRC [15] which uses the hand palm as a way to replace the RC.

The use of 2nd devices (cell phones and tablets) replacing RC is also studied. Several papers analyze these interactions as it is shown in [16], to control basic TV functions and selecting and sharing TV material. However, none of these proposals is directly focused on the input of data in text format.

The virtual keyboard and multi-tap keyboard allow for data entry using standard RC from TV devices. However, in certain cases, data entry is substantially jeopardized because the user has to maintain the RC position and is usually slow to select the desired letters to input short or long texts [13]. The multi-tap

keyboards have best performance than the virtual when it comes to short texts due to its popularity among users of cell phones with numerical keyboards. Research indicated that its usage allows for considerably accelerated text entry [17]. However, touching buttons multiple times while keeping the RC directly pointed at the TV does not favor ergonomics. The usage of these two types of RC does not allow for long interactions which demand a more versatile alternative for entering data.

The best results obtained in those studies were achieved with devices that respect the naturalness of the user, whether based on point-select modality, or making use of one's own body (and its movements) to send commands. These solutions, in part, also contribute to the user's attention allocation. But one can assume that there is physical fatigue in prolonged use, such as when the user has to handle text input or other advanced commands while keeping his/her arms raised, for example.

The present research is justified since none of the existing studies show an affordable solution for data input. In the next section, we present the context which motivated us to study some theories of attention.

IV. HUMAN ISSUES IN INTERACTION WITH iTV SYSTEM

Our perception of the divided attention problem led us to study the theories of attention, the tasks performed by the user when handling RC data input, and the cognitive processes and factors involved in those tasks.

Three theories of divided attention deserve attention [12]:

- **Capacitive Theory:** Argues that a limited pool of cognitive resources is available. It predicts that an individual, who suffers an increase in the number of attention targets, necessarily reduces his/her cognitive resources available for each of the targets. Attention targets shall be defined as actual or imaginary-symbolic elements that require active or passive attention;
- **Cross-talk Theory:** attributes errors and delays in multitasking to interferences between the contents of the information being processed. I.e.: A viewer who tries to follow an iTV program while talking about a different subject, will either suffer losses of program content, or have trouble following the conversation; and
- **Automaticity:** refers to either innate or learned activities, of which action is automatically performed. Automaticity has as behavioral parameters: fast response time; obligatory execution; no interaction with other concurrent processes; constant performance level, no matter what other processes run in parallel, and less sensitive to distractions. The importance of this theory to the interaction takes place through the natural little effort needed to accomplish the interactive task.

Roda [12] notes how divided attention often induces errors and delays in responses, and most attention researches focus on multitasking performance and in identifying its influencing factors.

Table I shows the number of 'targets' to divide the resources of cognition and the number of tasks which the user performs

using a RC for TV data input. Switching between multiple tasks and targets of attention make typing a tedious task, and it can hardly reach a state of automaticity (as in typing on a full QWERTY keyboard).

TABLE I - SWITCHING TASKS WHILE TYPING ON THE TV

Cognitive Target	The User Task
-	1 - Thinking about what he/she wants to text
Control Keys	2 - Seeing and understanding which RC key does what
Screen Cursor	3 - looking at the TV screen and seeking on which letter the cursor is
Desired Letter(s)	4 - Finding the desired letter in the TV screen, and seeing how far from the cursor the desired letter is
TV Screen	5 - Pointing the RC to the TV correctly and typing the letter
Text Box	6 - Checking if the letter was typed correctly

In order to minimize user fatigue in the process of inputting data, we sought to understand in which interactive situations the division of attention causes no loss of the individual's performance. Wickens [12] identifies types of resources that influence the tasks, and situations where interference is minimal. Here we cite three: perceptual modalities, code processing, and visual channels.

- **Perceptual modalities:** The studies predict that different sensory modalities (visual, auditory, somatic (touch), kinesthetic (joints and muscles), etc.) use different non-conflicting resources. This indicates that separating the activities by using different sensory interaction modalities in the same moment of interaction can be a determining requirement for performance gain and user attention in the prototype. Data input using movement – when the user moves his/her wrist to select letters – and receives haptic (relating to touch) vibrating feedback of the movement between letters leaves one's sense of sight free to remain fixed on the TV screen;
- **Code processing:** Predicts that analog and spatial processes use attention resources different from categorical and symbolic processes. The separation of these types of processes occurs, for example, with the use of movements (spatial) and visual symbols on the TV screen (alphabet, cursor, etc.). Separating the activity of selection and vision, letting the user select letters without looking at the RC, can be a determining factor for better performance; and
- **Visual channels:** Predicts that human focal vision uses resources attention different from those of peripheral vision (environment). One way to benefit from this parallel processing is to design the use of the remote control so that the viewer's visual focus is on the keyboard from the TV screen, but peripheral vision encompasses the RC held in the his/her hand, assisting in his/her spatial perception.

Attention-aware systems reduce the information overload, limit the negative effects of interruptions, increase the situations of human knowledge on the environment (awareness), and support the user in multitasking situations [12].

Another study we conducted on human factors was with regards to the Ergonomics of the RC for a more natural interaction. It is a consensus that TV data input activity that lasts a long time should be supported by solutions that seek to maintain the user's body in a relaxed and comfortable position, and this assertion is another determining factor for improving performance.

V. PROJECT METHODOLOGY

The proposed solution is based on theories of shared attention to minimize problems of task switching, and was designed through evolutionary prototyping, whereby we sought to minimize discomfort in interaction. The used methodology is the evaluation of prototypes considering users diversity [18].

The prototype was developed and refined in three versions (see Figure 2) using the design rationale [19] and an iterative design process, which involved three main activities: definition; functional and usability testing; and calibration of the prototype.

The initial requirement of the prototype was to provide support for the user to perform the following actions in a natural and relaxed manner: locate a symbol (letter, number, other), select a symbol, cancel a selected symbol, and view feedback.

During refinement, we worked to properly identify which factors (allocation of user tasks among the TV artifacts, positioning of buttons and feedback, hand and head movement, etc.) impacted user performance and satisfaction when carrying out these actions.

The purpose of version A was to capture motion by using a gyroscopic chip and a single select button, in order to record the movements of the wearer's wrist. This characteristic functions by dividing the character selection task on RC and the viewing task on TV (challenge 1). Implementation of this prototype enabled us to evaluate and calibrate the sensitivity of the gyro sensor for inputting data on the TV.

Version B works the feedback of user's actions (challenge 1) with adoption of a vibration motor (which confirms actions made through the RC) and a joystick-type control for inputting and handling data (featuring with audible and haptic feedback for character selection). The ergonomically shaped body of the RC, with tilted LCD screen and USB communication, was aimed at better ergonomics and eliminating the need to point to the RC at the TV (challenge 2). These new characteristics were implemented by observing the use of prototype A. Prototype version C features a touch sensor at the tip of the joystick to identify when the user merely touches the device, including a new form of interaction with the joystick: touching, in addition to moving and pressing. Users used to mistype using the B prototype because the act of pressing the button also moves the cursor from the desired letter. Thus, the addition of a touch sensor improves the accuracy of users' selection and the feedback of actions and comfort when selecting symbols (challenges 1 and 2).

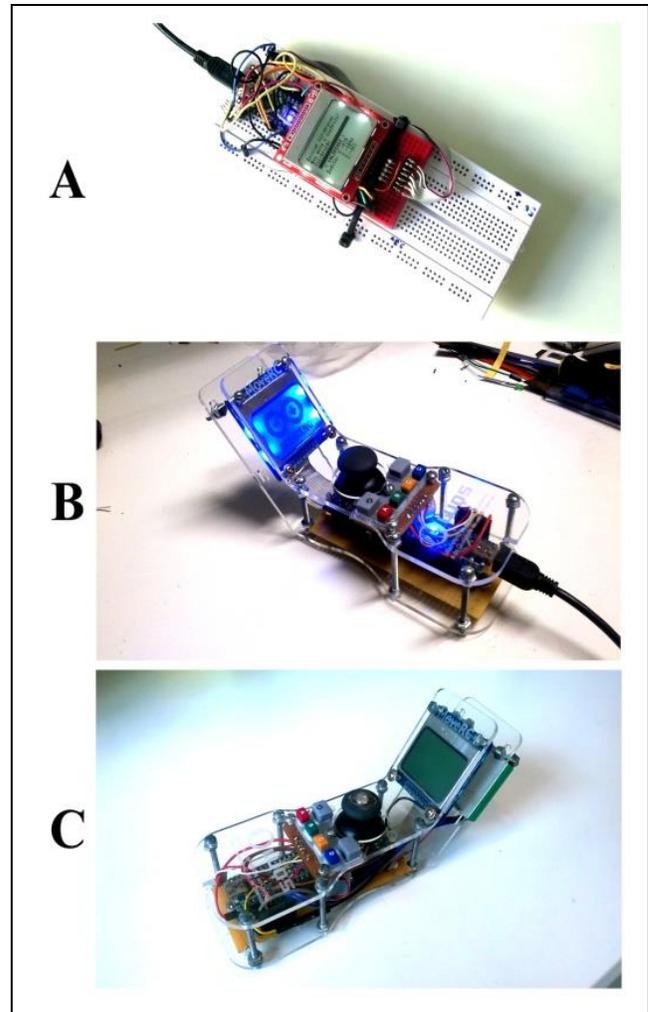


Fig. 2 - Three MoveRC prototypes

The requirements described in this paper are grouped in two: i) human performance: including the situations in which the division of attention between the user's actions should not cause loss of human performance; ii) user's satisfaction: including the requirements in which the interaction should be as natural as possible, such as the ergonomic and usability of the device, and the requirements in which the solution should have an affordable cost. The requirements are the following:

Requirement 1: Separating tasks in different sensory modalities. In this proposal the device allows two modes of interaction: simple interaction with simplified buttons, and advanced interaction (using motion capture). Data entry is accomplished through the use of movements. The user moves his/her wrist to select letters and receives a haptic feedback (vibratory) motion between letters, leaving the sense of sight free to remain fixed to the TV screen;

Requirement 2: Separating the viewer's perception of the selection and execution activities. This can be a determining factor to mitigate the sight alternations. To do so, the MoveRC prototype uses the gyroscope to identify the choice of letter by motion. The user does not need to lay his/her sight on the RC,

now that he/she will handle a single key, the one of letter selection. The single selection button also has a tactile feedback, so the user can easily identify if the button was pressed and if the command was sent to the TV;

Requirement 3: To keep the main visual focus of the user on the TV screen, letting his/her peripheral vision encompasses the RC. In the proposed solution, the viewer's visual focus is on the TV screen's virtual keyboard, and his/her peripheral vision reaches the RC on his/her hands, helping him/her with the spatial perception. The use of a screen on the RC can help people (with impaired vision or not) know which letter they are targeting;

Requirement 4: To assure automation. Two important aspects of the device are the following: i) the device provides immediate feedback of the actions, whether by sound, vibration or feeling a click when pressing buttons; and ii) the system response is immediate, with visual or haptic feedback;

Requirement 5: To be ergonomic. The aspects of the device for this requirement are the following: i) the pointing device is simple, respecting the standardization of television RCs; ii) the device uses another form of data communication instead of infrared (USB or Bluetooth), allowing the two-way transmission of data without worrying about the RC positions; and

Requirement 6: the interaction device has an affordable cost for diverse profiles of users who are buying a digital TV, sold with an InfraRed Remote Control (IR RC).

Specification and Implementation

For the choice of physical computing platform, the criteria were: ease of programming, low cost, ease of integration with sensors and other equipment, and prior knowledge of the participants. The hardware platform used was Arduino. In order to simplify use, the MoveRC maintained the recommendations of Carmichael [20], with the number of buttons minimized. The physical characteristics of the MoveRC are grouped in input and output specifications:

Inputs:

- Gyroscopic sensor, 3-axis sensing of movements made with the wrist;
 - Joystick sensor, 2-axis sensing of thumb movement;
 - Touch sensor, sensing the touch of the user's thumb on the joystick;
 - Joystick pressure button, sensing a pressing force stronger than touch;
 - 6 TV standard colored function buttons (back, menu, red, yellow, green and blue).
- Outputs:
- LCD display 84 x 48 pixels (vision) tilted at a 40° angle;
 - Vibrating motor (tactile and audible);
 - Joystick button click feedback (tactile and audible).

Interaction Language

Interaction language is the definition of expressive codes which users must use to communicate with the system, and is constructed by interaction designers [21]. The interaction language created is based on the user's wrist and thumb movements for performing the aforementioned user actions. Figure 3 shows the iTV application's help screen (created for the solution), and demonstrates how the user should handle the MoveRC.

In the steps below, we highlight the codes (commands) that will guide a situation of interaction during the user's movements:

1) *The user, without touching the analog joystick cursor, must move his/her wrist (similar to the movement that one makes when using a computer mouse) to move the on-screen circular cursor. This movement allows good speed for selecting letters, even ones that are far apart;*

2) *Upon locating the desired letter, the user should touch the analog joystick cursor. At this point, the gyroscope is turned off and the cursor changes from a circular shape ● to a square with arrows around ◼, indicating that at this time it's necessary to use the analog joystick button to move the cursor;*

3) *If the user finds the desired letter in step 2, suffice it to press harder on the analog cursor for the letter to appear in the text. If the user makes a wrong selection, he must simply move his/her thumb to make small horizontal and vertical movements to correct the selection, and then press the analog stick to select the right letter; and*

4) *To clear errors, the user presses the dedicated button to the left of the analog button.*

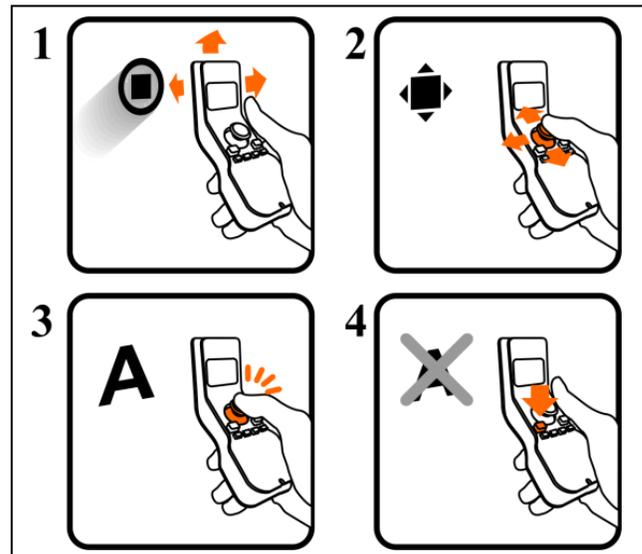


Fig. 3 - MoveRC Interaction Language

VI. EVALUATION OF PROTOTYPE

The hypotheses raised in this paper were evaluated in usability trials with version C of the MoveRC.

A. Evaluation ecosystem

An evaluation ecosystem was created, which allowed monitoring and data collection on the interactive actions of each user, as well as identification of the users patterns of attention and relaxation, while interacting with an iTV application for inputting data via a RC. The ecosystem is composed of an iTV application and the integrated solution, called USATT, to show collected data on usability and attention in an integrated way.

1) The iTV application and its use in digital TV via different remote controls

An iTV application was developed in this project to be run according to the digital TV standard used. The application was used for inputting data for use with an IR RC and with the MoveRC (Figure 4), for the purposes of comparison of user performance, satisfaction and patterns of attention and relaxation manifested thereby. It was implemented in NCL+Lua and Processing, and features a text box on the keyboard and a cursor according to the position of the RC being used.

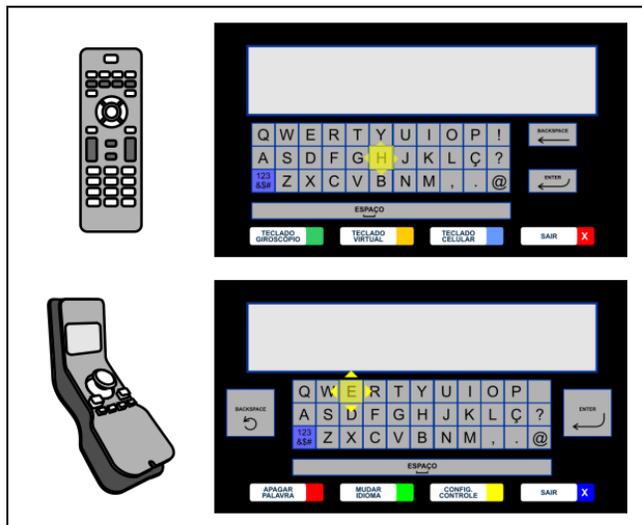


Fig. 4 - IR RC and iTV application (above) and MoveRC and iTV application (below)

2) USATT solution: Attention-capture device integrated with MORAE

The attention and meditation of a user is normally analyzed qualitatively, using scores (much, little) assigned to his/her behavior in interaction (such as: interested, distracted, bored, etc.). This solution is a subjective analysis, thus setting qualitative data. A different approach, called USATT (Usability-Attention integration), was used, with the adoption of a portable EEG device for measuring brain activity, providing quantitative data for analysis. We developed a way to integrate the information collected by this device with information collected by the usability test software, Morae, that captures data on the performance of users during interaction.

The device used in the tests was a NeuroSky Mindwave EEG headset, connected to a computer dedicated to recording the usability trials. The software used to the task was Mind

Stream [22]. According to Wróbel [23], the user’s Alpha (8-13 Hz), Beta (15-25 Hz) and Gamma (30-60 Hz) waves are responsible (respectively) for the states of meditation, attention and perception of the human mind. This device transmit information to the computer regarding the user’s attention (values from 0 to 100%) and the user’s meditation or relaxation (values from 0 to 100%), at about 40 readings per minute. Attention near 100% indicates that the user is focused on the task being performed, while attention close to zero indicates a user distracted from the task. Relaxation close to 100% indicates a user who is tranquil with the task being performed, while relaxation near zero indicates a user who is nervous or tense with the task. Neurosky Mindwave also sends a quality signal data (0 to 100% error reading), but no threshold was necessary to implement, because data with more than 30% reading error is not transmitted to the computer.

Data from usability test software (with starting and ending times for the tasks, and records of errors and details of the test) were merged data from the EEG. Both data were in a time stamped CSV format, witch permitted to synchronize the events. A software – implemented in Processing IDE especially for this study – transform the data from the EEG and usability test software in graphics, containing data on usability and attention in an integrated way. In the graph in Figure 5, the time spent by users is represented by the horizontal black line running left to right. The blue line and green line respectively record levels of attention and meditation of the user over time. Red, blue and yellow vertical lines indicate, respectively, moments when errors occur in the task (red) , user requests for help (blue) and user’s interjections (which denote failures in communicability system [21]).

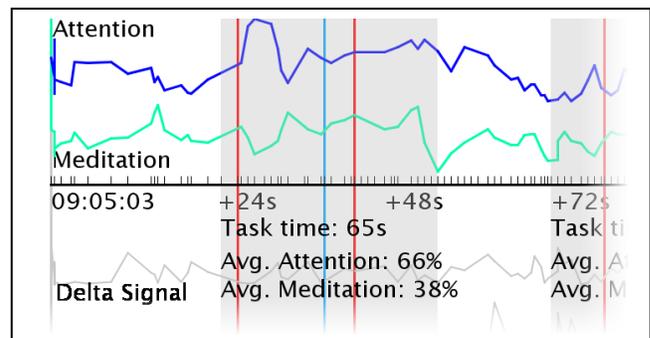


Fig. 5 - Usability and Attention Graphic Detail

The graph also shows a summary of the data on average attention, average meditation and execution time of each task. Attention and meditation data between tasks were not consider for this analysis.

Other signals from the Neurosky Mindwave were also shown in the graphics (horizontal gray lines below the black time line). It represents all other brain waves recorded by the EEG. But those data were not used for this analysis.

This kind of graphic allowed us to analyze the user’s reactions of attention and meditation during the tasks. For example, it is possible to evaluate attention and meditation of a user before and after an error occurrence.

B. Usability Testing

The tests were conducted in a controlled environment (soundproof room). The iTV application was transmitted along with the televised content to a TV located 1.5 m from the user. The TV video content utilized (which is outside the scope of this research) consisted of peaceful images of nature without audio, in order to focus the user’s attention only on the tasks of the usability test. For this research, two (2) pre-tests and eighteen (18) tests were conducted, in individual sessions, whereby the participating users had to perform scenarios, carrying out the actions described in language designed. Tasks included creating a login and password on the TV, and subsequent typing of a long text, simulating the situation in which a user has just purchased a TV and is prompted to register in the system.

Each participant used the infrared RC and MoveRC, alternately. Each session lasted 30 minutes, on average.

All tests were videotaped for subsequent analysis. Two cameras were used to record the tests: One recording the task performed on the TV screen, and the other recording the user’s facial and body expressions (see Figure 6).



Fig. 6 - Recording the usability test

In addition to the aforementioned automatic collection by the software, the users answered questions on the ease of performing the tasks. At the end of a test session, we used the System Usability Survey – SUS recording the user’s satisfaction with the solutions.

VII. ANALYSIS OF THE DATA

All of the users performed the usability tests with both RC solutions, and answered the questionnaires. Hence, we have paired data, whereby statistical analysis can be used to the test of the hypotheses, to look for the differences between data of user groups and to verify the implemented requirements.

A. Sample

Altogether, 18 participants were evaluated (not including two pre-tests). The mean age of 18 participants was 30.8 years and the gender divide was 33% women and 67% men. The education levels were: 11% high school diploma; 27% college

degree, 33% post-graduate certificate, and 27% master’s degree, all middle class. (details in Table II).

The data on task execution times follow a normal distribution, when applying the Shapiro-Wilk normality test ($W = 0.8725$, $p\text{-value} = 3.587e-08$). The same occurs with SUS data ($W = 0.9504$, $p\text{-value} = 0.1073$).

TABLE II - SAMPLE

Users	Gender	Education level
1	Male	College degree
2	Female	College degree
3	Female	Post-graduate certificate
4	Male	Post-graduate certificate
5	Male	Master’s degree
6	Male	Master’s degree
7	Male	High school diploma
8	Male	Post-graduate certificate
9	Male	College degree
10	Female	College degree
11	Male	High school diploma
12	Female	Post-graduate certificate
13	Male	Post-graduate certificate
14	Male	College degree
15	Male	Post-graduate certificate
16	Male	Master’s degree
17	Female	Master’s degree
18	Female	Master’s degree

B. User Satisfaction Analysis

Analyzing the boxplot (Figure 7), one can see that the SUS average for the MoveRC (80th percentile) was higher than that of the IR RC (69.58° percentile). With both values above average, it can be stated that the solutions are well accepted by the user, and that that are consistent with the usability standards followed by the industry.

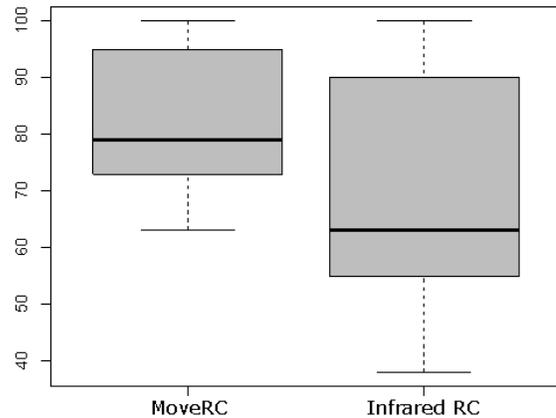


Fig. 7 - Boxplot SUS data

Even with an above-average value, the variation in the SUS index of the IR RC shows that this solution had lower indications of satisfaction compared with evidence collected using the MoveRC. The height of the boxplot (Figure 7) for the infrared RC makes it clear that the solution received scores ranging from 40 to 100, indicating heterogeneity of opinions on this form of interaction.

The variation of the MoveRC, aside from being lower, is positioned above the 60th percentile, indicating that the solution was generally better accepted by users. The average evaluations for the MoveRC equaled 80.2 points, and 69.7 for infrared RC, a difference of 15% more for the MoveRC.

Figure 8 shows the results of the questionnaire on the ease of performing the tasks, with index zero being difficult and index 5 being easy. Analyzing the data, one can see that for short typing tasks, the user understands that the IR RC is easier to use than the MoveRC. But as the typing volume increases, this ease of use of the IR RC decreases. Users reported that the IR RC was easy to use because the users already knew how it works. But for long texts, the task became more tiring. The MoveRC, on the other hand, maintained an average score of 4 (easy to use) for all of the tasks. The users reported that the once the initial difficulty in understanding the sensitivity of the motion capture was overcome, the task became less wearisome.

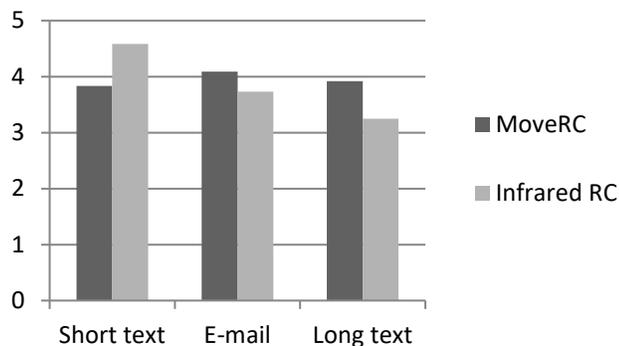


Fig. 8 - Easiness in the tasks

We performed Null Hypothesis Significance Testing using Student's T-test for paired data (paired test) of the SUS questionnaire from users in each solution. The null hypothesis is: The use of a device for data input that meets the requirements relating to user attention and ergonomics and does not interfere with user satisfaction in TV data input, when compared to a solution that does not take into account such requirements. The level of significance was $\alpha = 0.05$ (5%).

Student's T-test revealed $t = 2.637$ with 17 degrees of freedom (18 users - 1, for a paired test), p -value = 0.0173 and mean differences in data between groups of solutions = 10.4444. T-Student's one-tailed probability table gives a value of 2.110 for degree of freedom 17 and probability 0.025 (. Thus we can reject the null hypothesis and accept the alternative hypothesis, since $t = 2.637 > 2.110$.

C. User Performance Analysis

We compared the execution time of the tasks (Figure 9) for each test solution. The task of typing a short text showed that the typing time with the IR RC is shorter than with the MoveRC. In both solutions, the execution time of the task grows as the typing volume increases. But this growth is most evident in the IR RC. This is achieved by the number of strokes required on the RC to type a single letter.

The averages of the execution time data of the MoveRC increased very little, even with the increased volume of text typed. This is achieved by the user's learning to manipulate the control of movements during the tasks, as evidenced in the post-test reports. The users' comments indicated that additional time is needed to get used to the motion capture solution. This characteristic is in line with what is informally called the law of practice, which states that the reaction time of a task decreases linearly with the logarithm of the number of attempts [24], [25].

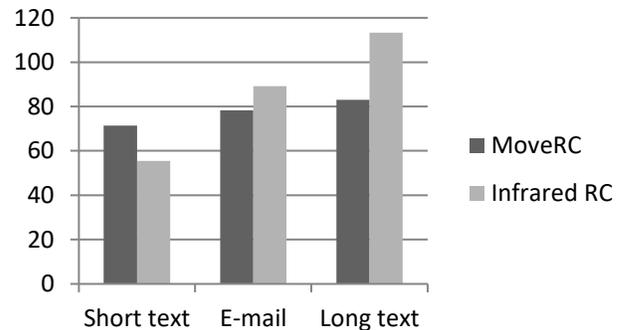


Fig. 9 - Execution time in tasks

For a more conclusive analysis of the data, Null Hypothesis Significance Testing was conducted using the Student's T-test (paired test) of the users' times in each solution. Student's T-test returned $t = -1.1229$, with 17 degrees of freedom and p -value = 0.2771. For a one-tailed probability of 0.025 with 17 degrees of freedom, we have the value of 2,110. Again we can reject the null hypothesis because: $t = -1.1229 < 2.110$.

D. Comparing between expert and non expert users of MoveRC

Since the average execution time from task 1, short text, was worse with MoveRC, a new evaluation was made as to understand this weak performance. Post-text interview mentioned that, it being the first contact with a device like that (which uses movement for selection) the users felt that if they had had more practice, they could better manage the typing task. A future evaluation could focus on the evolution of execution times of a novice user, applying the 'power law of practice' [24].

E. Attention and Meditation Analysis

Analyzing the graph in Figure 10, one can see that the indices of average user attention are similar in both interaction solutions, MoveRC and IR RC with virtual keyboard (54.7% and 53.8% respectively). However the data on meditation show that the MoveRC keeps users more relaxed during the tasks.

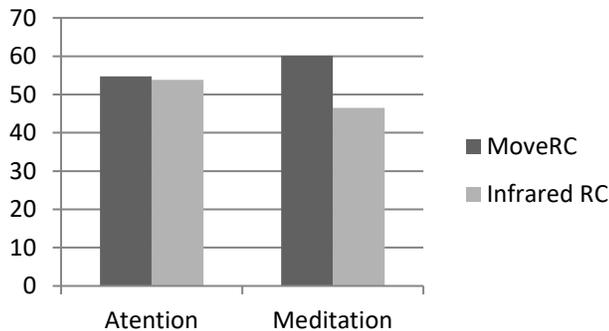


Fig. 10 – Average attention and meditation in the tasks

The average for meditation with the MoveRC was 60.2%, vis-à-vis 46.5% of the IR RC, an increase of 29.4% in average user relaxation. The fact that the MoveRC solution promotes higher performance in the tasks, increasing user relaxation and maintaining the same levels of user attention as a common RC may be regarded as evidence that the MoveRC better allocates the user's attention in the tasks.

F. Analysis of Requirements

This is an analysis described based on evidences perceived by the authors about the use of MoveRC during the three tasks demand for each one of the 18 users in usability test sessions.

This description is organized based on the three W3C principles for an interaction product project [26], also applied in Piccolo et al [27].

To analyze each principle, we identify the requirements that are connected to that principle. After that, we enumerate the investigative elements from the solutions (MoveRC and TV application user interface) implemented, and selected the most relevant and non-ambiguous. Then, we analyzed the human aspects (like emotional, human effort, potentially for improvement, etc.) from the users, looking for the selected elements. We made use of the video recording sections, to capture corporal expressions and their speeches, as well as data from interview, checklist, usability test software data; lining up, when necessary, this analysis with other already done.

The W3C principles used are:

- **Understandable:** Information and the operation of the user interface must be understandable [26]. This item refers to interaction modalities, vocabulary and metaphors used in the solutions implemented for requisite 1.
- **Perceivable:** Information and user interface components must be presentable to users in ways they can perceive [26]. We consider in this principle the RC design decisions which resulted in elements perceived and not perceived by the users. This item refers to requisites 2 (Does the user perceive a letter in navigation, to choose and choose) and 4 (Does the user perceived/feel the feedback (click sensation, vibration, sound) as much in application as in the RC?).
- **Operable -** User interface components and navigation must be operable [26]. Human aspects are included in

this item, such as the user's effort and his/her potentialities to become an expert when interacting with the TV (such as when inputting data and using the RC). Those are affected by the ergonomics of the Solution (requisite 5).

First we verified if the interaction solution (movements and joystick button) to navigate and to select actions were understandable by users. We also observed their comments to identify evidences about such understanding.

Results identified the users hardship in understanding the sensitivity of the motion capture, especially in the beginning of test.

We searched to know if:

- Users understand how to insert and fix a letter, i.e., how to hold and release the analog joystick to select an item?
- Users understand when one option can be selected, i.e., does s/he understand on-screen cursor metaphor (the cursor changes from a circle to a square with arrows).

The user's speech and her/his reaction analysis at typing a text revealed that the metaphor of the cursor/joystick was understood. Two user comments referring the selection solution are highlighted: User 6 commented during the realization of the Task 3 (T3). "We don't need to put the RC exactly over the space bar, it (the remote) already feels our intention of using the space". User 4 understood the solution of moving the wrist (with the RC) to select the desired letter, when finalized T2. He said: "Its sensor (referring to the RC) is not on the TV; I mean, we don't need to point the remote to the TV to select a letter".

Users also understood the metaphor of on-screen square cursor (Upon locating the desired letter, the user should touch the analog joystick cursor. If the user finds the desired letter in last step, suffice it to press harder on the analog cursor for the letter to appear in the text). This can be proved considering the quantity of wrong letter selections that were fixed (pre-selected) before typing. So, we investigate the user selection accuracy. A prudent user (with an average performance of T1 = 70 sec.) made 7 to 9 wrong letter selections, and typed at most 1 wrong letter. A high performance user (T1 = 43 sec.) pre-selected 5 wrong letters by task, and didn't type any wrong letter. Those who used the re-selected function a lot (users 2 and 18) talked that the solution was too sensitive, referring the fact that the on-screen cursor changed fast to the selective state.

In the second item, we evaluate how users perceive/feel the feedback from the RC.

- Does the user feel the vibrating and clicking feedback?
- Does the user perceive de visual feedback on TV screen?
- Does the user perceive the visual feedback on the second screen?

The vibrating feedback was little commented. User 6 commented in the beginning of T1 about this sensation: "It's quite sensitive; when it feels that we put the finger next to the button, is like it already selects."

This user, after having made the two initial tasks moving the wrist naturally, began to do inaccurate circular movements with the RC, in an attempt to do the selection. Then he felt the click sensation, understanding something was wrong. So he commented his feeling, and the evaluator warned him: "When you touch the finger over the button, it locks the selection". The same problem happened with user 7, who commented: "It is locked, do you know?" User 18 also complained: "Sometimes this RC doesn't work".

It is important to say that some end results showed one or more space errors (words without space between them, or too far). Among 18 users, 4 of them (users 3, 4, 10 and 15) wrote their texts putting two spaces together. Two (users 4 and 10) didn't type spaces between words. Some users related this issue saying, "I thought that I had already typed the space". This problem may have happened because the difference of perception of square cursor over the wide space bar (some users complained about not being sure if the cursor was over the space bar).

On the second screen we observed that some users (3 of them) interacted alternating their observation between the TV and RC. This situation happened with a user who was not using his prescription eyeglasses (user 4) and with other two users, who, for a short time, were confused on how they should manipulate the RC (users 5 and 18).

Those results on user reactions through typing revealed that all users were aware about the visual and sensorial feedbacks. All users could generate results (words or phrases) they wished, preventing and fixing errors when was convenient for them. The second screen was less perceived than expected, however satisfactory. Thus the proposed requirement (to help users with or without sight problems to select a letter in the virtual keyboard) was validated.

According to the Operable principle, our evaluation focus was on an ergonomic solution. In videos, 14 users used the MoveRC as predicted (they rest the arm above their thigh, moving wrist only). Three users insisted in using the MoveRC with both hands (similar to a videogame user manipulating a joystick). One of the users handled MoveRC with the extended arm without support, doing rough movements in the air (but only in part of the test). User 4 was the only one to comment on the moving wrist solution (with the RC) to select, as cited previously. Our analysis revealed that most users were aware about the MoveRC operability, and naturally found which position was most comfortable to use it.

VIII. DISCUSSION

A. *Implication for Specification and Evaluation of Interaction Aspects for the iTV Systems*

In this sub-section we present the contribution of this work to the specification of interactive solutions for iTV, as well as for validating solutions.

1. *Use of physical prototyping technique*

We applied the physical prototyping technique to implement the specification of an interactive solution for iTV, subject to refinements.

A physical prototype is a promising feature, as it allows developers to design, implement and test interaction solutions without having to involve other agents responsible for the production of physical equipment (such as TV and RC manufacturers).

In addition if we had chosen an alternative system of low-fidelity prototyping, such as the Wizard of Oz, a paper prototype, we could not have seen the economic viability of the solution. An IR RC and a smart RC cost approximately US\$ 10 and US\$ 100, respectively. The final prototype of MoveRC to mass production has a cost of just over US\$ 25, given the simplicity of its circuit and the low number of sensors. It is important to point out the interaction device must have a cost that permits to be acquired together with the TV set. Edge solutions, like movement capture with Kinect, or those who need complex installations on the iTV environment, are not economically viable.

Another feature presented here refers to the specification of the command language, which was described by associating digital aspects (such as the activities and forms of interaction) to prototyped physical aspects (such as buttons, RC sensors). This specification ensures the reuse of solutions, while keeping the decoupling between them. It makes it possible, for example, changing a physical solution without necessarily having implication for the digital solution.

2. *Evolution of requirements*

Our proposal considered six requirements, which were observed as they evolved (had new features added) and new requirements defined. In this section three new requirements are described, followed by four complementary features to the proposed requirements:

- The use of text prediction dictionary is recommended for long text entry;
- The interaction device must respect the accessibility rules and guides, like: to have tactile marks in buttons, and buttons with different functions in different shapes and forms;
- Voice recognition must permit intuitive commands, like 'turn on TV' and 'TV on'. To text entry, it should use an updateable dictionary, containing regional terms and slangs. It's important to turn down the TV volume during the capture, to keep the user auditive attention in her/his own speech;
- The interaction must be as natural as possible, using human inner actions. It should not be designed in a restrictive form (with movements and commands that must be recorded). Another solution to be considered is eye movement;

- The interaction device must use other form of transmission aside from infrared beams. It must provide bidirectional communication and fault prevention due to lack of line of sight between RC and TV. Possible solutions are radio waves or wi-fi;
- The use of a full QWERTY keyboard in the interaction device can solve, as long as the button layout is maintained clean, without many buttons together. Commercial solutions uses the opposite side of the remote, or collapsible doors;
- The interaction device must provide users with immediate feedback for the actions, by keeping them aware if a command was received by the iTV or not. Sound, tactile (in buttons) and haptic (vibration) are solutions for that.

The investigation of new requirements must be planned and can be done in different contexts (as in the laboratory, in simulated and real contexts) [28]. It is important to show stakeholders potential challenges of the interaction with iTV but also to bring their attention to good practices (see Figure 11).

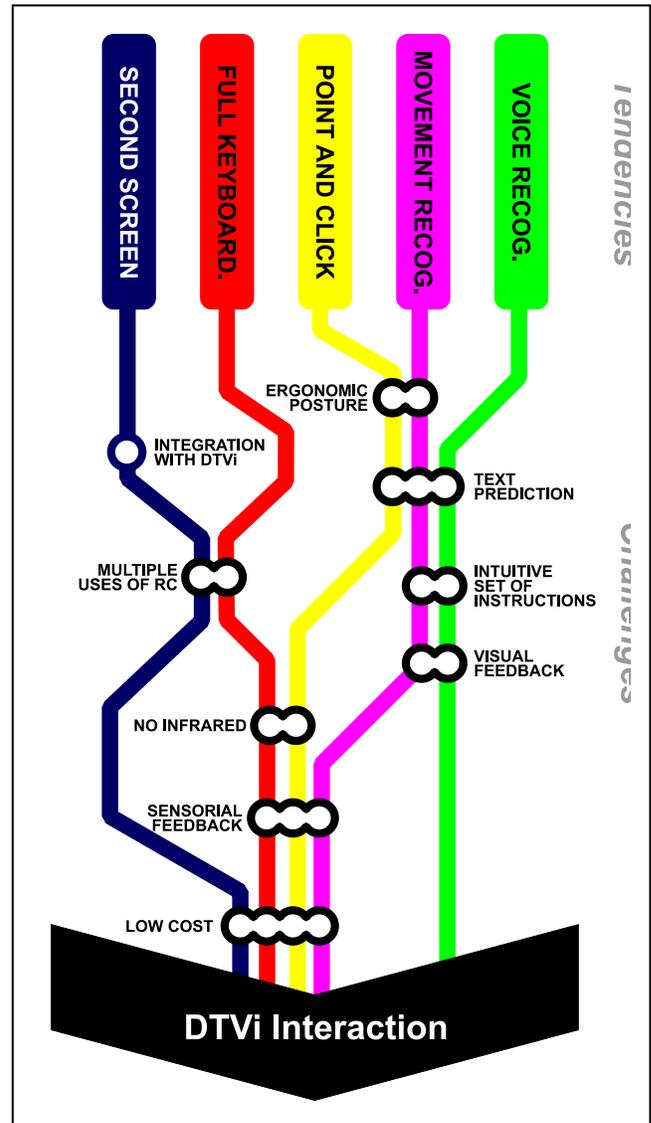


Fig. 11 – Tendencies and challenges to interact with iTV

3. Contribution for several agents involved with the iTV Systems

We describe, as follows, the advantages of this research in encouraging innovation, creation of patents, education and interaction between industry and Academia. They are:

- To encourage manufacturers of new devices to review their development and production process and introduce human aspects, in order to improve user experiences. They should also realize the importance of assessing the experiences of users from the beginning of a process for innovation and realize how to satisfy the customer, without focusing solely on the technological aspects and market;
- To create opportunities for industry and academia to work in partnership, associating technologically sophisticated processes with experiments and theoretical basis in order to create products that can be evaluated through hands-on experiments;

- To expand the requisites demanded from interaction designers who use ISO 9241-9 on data entry from devices that are not keyboards; and,
- To integrate undergraduate and graduate classes that work with physical prototypes and modeling (such as mechatronics, automation and electric engineering), to produce multidisciplinary projects of devices which present good usability with exclusive focus on technology, and supported by user studies.

B. Contribution for the future of brain computer interaction

The measure of human brain waves offers new perspectives for human computer interaction, such as new quantitative index to compare studies [29] or mind control of software [30].

The possibility of altering the conventional dynamic of usability testing becomes clear due to the importance of theories of shared attention in a dynamic context. Today, no application (mobile, web, IoT) can guarantee calm, silent and controlled scenarios. Gadgets are being used on the street, while walking, driving, during social interactions or as a second screen. This results in user attention becoming more fragmented between the environment (scenario) and our applications and devices. Another promising method of conducting usability tests is to quantify data about attention and thus decreasing subjectivity in the relation between the user and levels of difficulty and attention.

The quantitative measurement of attention and meditation, by means of EEG equipment, generates new possibilities for usability evaluation, to the extent that it quantifies data otherwise only qualitative in other studies.

EEG equipment has the potential to become as important as eye tracking systems, in test environments.

The collection of data directly from users' brains signals new horizons for human-computer interaction, not only in the collection data for usability but also in the very usability of systems controlled by the brain [30]. Future studies point to the use of this data collection approach in usability tests of other modalities of interaction with the iTV.

IX. CONCLUSION

The motivation for this research came from studies of new technological solutions to interact with the TV, which implement different modalities based on gestures, voice recognition, recognition of body movements, use of additional artifacts, etc., and of common remote controls used for inputting data in text format on the TV. Problems related to human factors (discomfort and cognitive effort) were identified. The analysis and design of a new RC solution was conducted in light of the theories on shared attention in order to minimize problems of task switching, and performance. The necessary requirements were implemented and evaluated.

The main feature of the proposed solution was to input data by movements of the user's wrist and the thumb, combined with a screen and auxiliary two-way communication between the RC and the TV through USB or Bluetooth technology.

The comparison of performance between the MoveRC and the IR RC took into account the execution time for tasks of typing short texts, email and long text. The statistical result of the execution time showed that the prototype MoveRC had a shorter time in all of the tasks proposed, and the time for typing long texts was reduced by 27% with the MoveRC.

User satisfaction was evaluated taking into account SUS data, together with the questionnaire on ease of the task. Both of the RC solutions had a mean SUS index above 68%, a value considered above-average, indicating that both solutions had a good evaluation from users. But the MoveRC showed a 15% higher satisfaction level measured by the SUS questionnaire.

Data collected on attention during MoveRC and IR RC testing showed that both solutions had similar attention loads for the proposed tasks. However, the indices of meditation, which measure the user's relaxation during the tasks, show that the MoveRC provided a 29.4% increase.

A joint analysis of these numbers with the results of performance and satisfaction presents evidence that the MoveRC provided a better allocation of the user's attention, because it maintains the same attention indices as the IR RC, and obtains improvements in the user's performance, satisfaction and meditation.

With the conclusions of the second hypotheses, and given the fact that performance, satisfaction and attention allocation are linked to interaction, it can be inferred that the main hypothesis is confirmed: The use of a device for data input that meets the requirements regarding user attention and ergonomics improves the user's interaction in inputting text on the TV. As future work, we suggest studying the interference of the external environment on attention during interaction, as well as visual and auditory interference of television programming on interaction.

The Arduino platform proved to be ideal, offering low-cost prototyping and manufacturing; it allowed the solutions studied to be refined when validating the requirements.

Finally, this proposal may help to motivate the creation of intelligent RC solutions equipped with sensors and interaction language that makes it easier for the user to operate the iTV system in its new forms of use: interactive, social and collaborative.

REFERENCES

- [1] S. Bachmayer, A. Lugmayr, and G. Kotsis, "New social & collaborative interactive TV program formats," *Proc. MoMM'09*, p. 121, 2009.
- [2] A. Carmichael, M. Rice, D. Sloan, and P. Gregor, "Digital switchover or digital divide: a prognosis for usable and accessible interactive digital television in the UK," *Univers. Access Inf. Soc.*, vol. 4, no. 4, pp. 400-416, Dec. 2005.
- [3] E. E. M. Huang, G. Harboe, J. Tullio, A. Novak, N. Massey, C. J. Metcalf, and G. Romano, "Of social television comes home: a field study of communication choices and practices in tv-based text and voice chat," *Proc. CHI'09*, pp. 585-594, 2009.

- [4] W. Cooper, "The interactive television user experience so far," *Proc. 1st Int. Conf. Des. Interact. user Exp. TV video*, pp. 133–142, 2008.
- [5] D. Geerts, "Comparing voice chat and text chat in a communication tool for interactive television," *Proc. 4th Nord.*, no. October, pp. 461–464, 2006.
- [6] S.-H. Kim, J. Ok, H. J. Kang, M.-C. Kim, and M. Kim, "An Interaction and Product Design of Gesture Based TV Remote Control," in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 2004, p. 1548.
- [7] L. Vignaroli, R. Del Pero, and F. Negro, "Personalized newscasts and social networks: a prototype built over a flexible integration model," *Proc. 21st WWW'12 Companion*, pp. 433–436, 2012.
- [8] T. H. Song, S. M. Jeong, M. K. Kim, K. H. Kwon, and J. W. Jeon, "Sensible interface using multi-sensors in mobile device," *Proc. 7th Int. Conf. Adv. Mob. Comput. Multimed. - MoMM '09*, p. 637, 2009.
- [9] L. C. De Miranda, H. H. Hornung, and M. C. C. Baranauskas, "Adjustable interactive rings for iDTV," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, 2010.
- [10] D. A. Chagas and E. S. Furtado, "MoveRC," in *Proceedings of the 19th Brazilian symposium on Multimedia and the web - WebMedia '13*, 2013, pp. 277–280.
- [11] D. A. Chagas, E. S. Furtado, and J. Nobre Jr., "Análise de Alternativas de Design de Mapas para TV Digital Brasileira Baseada em Multicritérios," *Simpósio Bras. sobre Fatores Humanos em Sist. Comput.*, 2012.
- [12] C. Roda, "Human attention and its implications for human-computer interaction," in *Human Attention in Digital Environments*, Cambridge: Cambridge University Press, 2011, p. 322.
- [13] L. Hoste, B. Dumas, and B. Signer, "SpeeG: A Multimodal Speech and Gesture-based Text Input Solution," *Proc. Int. Work. Conf. Adv. Vis. Interfaces*, pp. 156–163, 2012.
- [14] V. McArthur, S. J. Castellucci, and I. S. MacKenzie, "An empirical comparison of 'wiimote' gun attachments for pointing tasks," *Proc. 1st EICS'09*, p. 203, 2009.
- [15] N. Dezfuli, F. B. Müller, M. Khalilbeigi, J. Huber, and M. Mühlhäuser, "PalmRC: Imaginary Palm-based Remote Control for Eyes-free Television Interaction," in *Proceedings of the 10th EuroITV'12*, 2012, pp. 27–34.
- [16] E. D'heer, C. Courtois, and S. Paulussen, "Everyday life in (front of) the screen: the consumption of multiple screen technologies in the living room context," *Proc. 10th EuroITV'12*, pp. 195–198, 2012.
- [17] M. Ingmarsson, D. Dinka, and S. Zhai, "TNT - A Numeric Keypad Based Text Input Method," in *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 2004, vol. 6, no. 1, pp. 639–646.
- [18] L. S. G. Piccolo, M. C. C. Baranauskas, and E. Furtado, "A t-gov prototype evaluation considering a wide diversity of users," in *MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Interfaces and Human Computer Interaction 2008*, 2008, pp. 27–34.
- [19] J. E. Burge and D. C. Brown, "Integrating Design Rationale with a Process Model," *Work. Des. Process Model. Artif. Intell. Des. '02*, 2002.
- [20] A. Carmichael, "Style guide for the design of interactive television services for elderly viewers," Manchester, 1999.
- [21] S. Barbosa and B. Silva, "Métodos de Avaliação de IHC," in *Interação Humano-Computador*, Rio de Janeiro: Elsevier, 2010, p. 384.
- [22] E. Blue, "Mind Strem." Los Angeles, 2011.
- [23] A. Wróbel, "Beta activity: a carrier for visual attention.," *Acta Neurobiol. Exp. (Wars.)*, vol. 60, pp. 247–260, 2000.
- [24] A. Newell and P. S. Rosenbloom, "Mechanisms of Skill Acquisition and the Law of Practice," in *The Soar Papers (Vol. 1)*, P. S. Rosenbloom, J. E. Laird, and A. Newell, Eds. Cambridge, MA, USA: MIT Press, 1993, pp. 81–135.
- [25] S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, no. 1. 1983.
- [26] B. Caldwell, M. Cooper, L. G. Reid, and G. Vanderheiden, "Web Content Accessibility Guidelines (WCAG) 2.0," *W3C*, no. May 1999, p. 33, 2008.
- [27] L. S. G. Piccolo, M. C. C. Baranauskas, and M. E. S. Furtado, "A T-gov Prototype Evaluation Considering a Wide Diversity of Users," *Proc. IADIS Interfaces Hum. Comput. Interact.*, 2008.
- [28] E. Furtado, F. Carvalho, A. Schilling, and D. Falcão, "Projeto de interfaces de usuário para a televisão digital brasileira," in *Simpósio Brasileiro de Interação Humano Computador - IHC'05*, 2005.
- [29] R. Folgieri, C. Lucchiari, and D. Marini, "Analysis of brain activity and response to colour stimuli during learning tasks: an EEG study," *Proc. SPIE*, vol. 8652, p. 86520I–86520I–12, 2013.
- [30] M. A. L. Nicolelis and M. A. Lebedev, "Principles of neural ensemble physiology underlying the operation of brain-machine interfaces.," *Nat. Rev. Neurosci.*, vol. 10, pp. 530–540, 2009.